REMEMBERING YOU READ “DOCTORAL DISSERTATION”: PHRASE FREQUENCY EFFECTS IN RECALL AND RECOGNITION MEMORY

BY

CASSANDRA L. JACOBS

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Professor Gary S. Dell, Chair
Professor Kara D. Federmeier
Professor Cynthia L. Fisher
Professor Aaron S. Benjamin
Associate Professor Lili Sahakyan

ABSTRACT

Speakers understand and produce common words like *cat* more easily than less common words like *panther*. Similarly, this pattern of behavior shows up at larger levels, processing common combinations of words like *alcoholic beverages* more quickly than less common ones like *psychic nephew*. As a result, many researchers have concluded that these combinations of words have word-like representations in long-term memory as a way of explaining how both words and phrases can be easier to process the more common they are. This dissertation challenges these assumptions by using episodic memory tasks such as yes-no recognition and immediate free recall of combinations of words, under the premise that word-like representations for phrases should lead to word-like patterns of episodic memory. The results and a corresponding verbal model demonstrate that combinations of words are processed more easily not because phrases have the same structures as words, but because of the strength of association between the two words within a phrase, which leads to facilitated processing.

ACKNOWLEDGEMENTS

I got to this point by doing a lot of things I probably shouldn't have. I have had the enormous privilege of telling people what I want to do and doing it – rather than being told or – god forbid – having to ask – and as a result have really exploited the kindness of others. I have gotten to live a really happy life in grad school – I wrote up my first year project in a basement in Pittsburgh the summer I lived there for a couple months to learn Estonian. I wrote my Master's thesis in Oregon while on a road trip with my best friend. I studied for my qualifying in a tent, and spent most of the summer of my fourth year on the road in Michigan before delving more deeply into Computer Science. I was almost stubborn enough to go to France for a semester, but moved to California for a year instead, where I sacrificed stouts for IPAs in exchange for no snow in the winter.

I am thankful first and foremost to my advisor, Gary Dell, for letting me roam the country all too often to clear my head, and for not minding coming into meetings too early. Gary has been a tremendous resource and knows something about almost everything. He's never stopped being supportive of me and has been more patient than I could have possibly asked for. More importantly, Gary has really helped me hone my senses of "why not" and "oh, that's why not." It's easy to have a ton of crazy ideas, but with Gary it's easy to make crazy ideas *work*.

My parents, Sandra and Mike, are two major and independent components in my being here. My mom worked her ass off when I was growing up and never made school not an option for me – she finished college the week before I did and she'll never understand how proud I am of her for it. My father is the kind of guy who licks rocks and makes jokes about geological time, and his skepticism about the ivory tower has helped keep me, uh, grounded. They only lamented to me about me not becoming an engineer until after I'd already decided to major in linguistics.

Between the two of them I had a lot of support in delving into somewhat obscure problems and without their support an academic career really would not have been on my radar.

I thank my sister Nicole, her husband, and her two lovely boys Rowen and Peyton, who inspire me every day to make the future better for the people who will be living in it.

I'm thankful to my friends Brendan, Ryan, Helen, Josh, and Kevin, who facilitated a ton of trips and periods of decompression. Friends made it possible to drive to and from Texas every winter, to hike the Lost Coast and the Superior Hiking Trail, be rescued from exurban Milwaukee after my car broke down on the way home from backpacking Isle Royale, to visit Grand Teton National Park over New Year's and not see a day above 0ºF. Friends made it possible to hit almost all the national parks during the move to California. They made it possible to appreciate the good things about California while listening to all of my anxieties about work and finding a fulfilling career. Because of them, I got to try a lot of good beer and eventually forgave Champaign for existing.

I'm thankful to the entire department of Psychology as I knew it – I got to study there during the golden years of psycholinguistics at Illinois. It seemed like for almost any problem there was someone was working on it. I am especially grateful to my collaborators and the members of my committee, who have helped shape the way I approach problems and have given me tools to think about problems more critically. The lack of dogma in the Cognitive division is something I'll probably chase after forever. Finally, I am thankful to the undergrads who worked with me and have now gone on to do great things – Abby, Amanda, Alissa, Brianna, Claire, Erum, Eve, Faustyna, Francis, Jace, Kammie, Kathy, Kristie, Marie, Maya, Morgan, Nathaniel, Nikki, Rachel, Sarah, Siggi, and Stacy!

TABLE OF CONTENTS

INTRODUCTION

Processing language requires retrieving linguistic representations from long-term memory, either to understand others or to produce an utterance. Likewise, many episodic memory tasks are also language tasks when they use linguistic materials as stimuli. A number of studies in the memory and language processing literatures have proposed mechanisms for how individual words are recognized, recalled, produced or comprehended. Similarly, combinations of words, or phrases, have received attention in both language processing and memory domains. In this dissertation, I will be comparing memory for individual words like *cat* with memory for adjective-noun phrases like *handsome wizard* as a function of frequency, which is known to affect both language processing and memory tasks. Of particular interest is whether the memory tasks identify similar mechanisms for the representation and processing of single words and meaningful multi-word phrases. Different outcomes in memory tasks involving phrases as opposed to individual words suggest different representations for phrases and words in long-term memory. Such findings could help resolve some long-standing conflicts in psycholinguistics on the comprehension and production of combinations of words, in particular the questions concerning the extent to which phrasal representations contain or are composed of word representations as well as whether linguistic abstractions such as the word CAT, are stored as such, or emerge from the experience of episodes containing the word CAT.

In Chapter 1, I will present results of a published study looking at the effects of the frequencies of multiword phrases on recognition memory. In Chapter 2, I investigate similar phenomena but in a phrase free recall task, rather than recognition. In Chapter 3, I outline a verbal model that accounts for effects of concreteness, word and phrase frequency on the recognition and recall of words and phrases.

CHAPTER 1: RECOGNITION

Researchers have carried out thousands of experiments in which word frequency is manipulated with the goal of understanding how words are processed, produced, and remembered. For the most part, this research demonstrates that low frequency words are less easily acquired, comprehended, and produced than more common words (see Ellis, 2002 for a complete review). More recently, the question of whether multi-word sequences (or phrases) might exhibit frequency effects has been assessed. As with common words, high-frequency phrases are associated with benefits in reading time (Bannard, 2006; Smith & Levy, 2013), phrase decision reaction time (Arnon & Snider, 2010), greater fluency and speed of production (Bannard & Matthews, 2008; Arnon & Priva, 2013; Janssen & Barber, 2012) and recall memory (Tremblay & Baayen, 2010).

Phrase frequency effects are of interest because they tell us about the cognitive mechanisms implicated in the production and comprehension of word sequences. The findings cited above indicate that the combination matters. A phrase is not just a list of words. More importantly, these results are analogous to the discovery in morphology that people are sensitive to the frequency of whole words, and the inference that word processing involves some knowledge of the whole as well as of the component morphemes (Bien, Levelt, & Baayen, 2005). However, there remain many questions about the exact nature of the mechanisms involved. There are two main issues that arise: compositionality and abstraction. In this chapter, I present five recognition memory experiments that address these issues.

Compositional or holistic representations

The compositionality issue concerns two primary issues. First, there is the representation of a phrase, by which we loosely mean the mental/neural codes implicated in producing and understanding it, and second, whether these codes are a predictable superset of the representational spaces involved in the production and comprehension of its parts. So, a person's knowledge of the phrase *red house* may be compositional, derived solely from their knowledge of its component words, *red* and *house*. If the phrase is not compositional, but instead holistic, a language user's representation of it might be largely separate from their representation of the component words.

Phrases vary in the extent to which their meaning is predictable from their parts, with the meaning of *red house* being much more predictable than the meaning of *red herring*. A phrase with an unpredictable meaning therefore may seem to require a largely disjoint representation. It is also plausible that such representations might also be employed for more predictable phrases as well. Indeed, the discovery of phrase-frequency effects has occasionally been taken to indicate that the representation of phrases is holistic. However, while such results indicate that speakers do encode knowledge of the sequences, they do not address the question of whether combination-specific knowledge is utilized instead of or in addition to word knowledge when processing phrases.

Episodic or abstract representations

The issue of abstraction concerns how we encode multiple instances of the same phrase. A phrase could be represented either as a collection of episodic memories, each containing a token of that phrase, or as a single abstract encoding of the type with an associated strength. In the episodic approach, the particular episodes in which a phrase is experienced are kept distinct, and effects of the phrase frequency would be attributed to the number of such episodes. In particular, any processing benefits that accrue to common phrases would be attributed to the greater availability of relevant memories to guide the processing (e.g. Goldinger, 2004;

Hintzman, 1988). Alternatively, in the abstractionist approach, each phrase type is a single representation such as a node in a lexical-semantic network (e.g. MacKay, 1982). If *red house* had been experienced a number of times, a node would represent the phrase type, with its strength (e.g. resting level of activation) proportional to its frequency. Of course, the abstractionist approach does not deny the existence of episodic knowledge about phrases. It simply assumes that the abstraction exists in addition to episodic memories, and it is this abstraction that plays the major role in how the phrase is processed, rather than the episodes.

Some accounts of word and phrase frequency effects are neither clearly episodic nor explicitly abstractionist in the sense that they have a single node for each word or phrase. Multi-level connectionist models (e.g. Seidenberg & McClelland, 1989) occupy an interesting middle ground in this respect. Each experience changes the weights in a network (as with an episode) and yet these alterations are not stored separately, but rather are superimposed. The resulting superposition is somewhat like an abstraction, but it is not easily recognized as such and is certainly not a single node. A related class of models, the naive discrimination learning models (e.g. Baayen et al., 2011; Baayen et al., 2013), also lacks discrete episodes and explicit representations of abstract items. For example, one such model by Baayen et al. (2011) consists of an input layer of letters and letter pairs and an output layer of semantic features. The model learns input-output mapping for words or phrases by applying the Rescorla-Wagner (1972) equations to probabilistic information obtained from corpora. Even though it lacks explicit words or phrases, its behavior (e.g. mapping accuracy) reflects both word and phrase frequency.

As we noted, benefits for high frequency phrases have been clearly demonstrated in comprehension and production tasks, and in memory recall. In our studies, we turn to a different memory task in order to address phrase frequency from a new perspective: the yes-no recognition task. Importantly, the high frequency advantage apparent in linguistic tasks and recall is not evident in recognition memory; in fact, low frequency words are recognized better. We more easily pick out *panther* when it was studied and reject it when it was not studied than a higher frequency word like *cat*. That is, low frequency words attract more hits and fewer false alarms than high frequency words. This pair of results is one manifestation of a broader category of what are called mirror effects, effects in which a particular class of items or condition of study for a set of items leads to them being more easily discriminated (Glanzer & Adams, 1985). The mirror effect allows us to derive predictions about frequency effects for phrases in recognition, and thus examine the cognitive mechanisms implicated in their processing.

In the next section we review studies of word frequency in language processing and acquisition. Next, we discuss the degree to which the high frequency word advantage is reflected in larger sequences of linguistic units, such as multi-word sequences. Finally, we review the mirror effect in yes-no recognition memory and consider its implications for multi-word sequences.

The high frequency word advantage

High frequency words are easier to process than low frequency words. The language processing system is adaptive and thus learns to process more probable events with greater facility (Jusczyk, 1997; Saffran et al., 1996; Lively, Pisoni, & Goldinger, 1994; Forster & Chambers, 1973; Dell & Jacobs, 2015). For example, identification of high frequency words is more robust under both noisy (Howes, 1957) and clear (e.g. Foster & Chambers, 1973) conditions.

When reading words in text, reading times scale inversely with the logarithmic frequency of the word that is being read, with the most common words in the language being barely read at all or even skipped entirely (e.g. Demberg & Keller, 2008; Howes & Solomon, 1951; Rayner, 1998; Smith & Levy, 2013). When a text contains low frequency words, comprehension suffers (Diana & Reder, 2006; Marks et al., 1974; Freebody & Anderson, 1983). In production, uncommon words are retrieved more slowly during picture naming (e.g. Oldfield & Wingfield, 1965) and produced less accurately (Dell, 1990; Kittredge et al., 2008). In short, the deck is stacked against low frequency words in linguistic tasks.

High frequency words are also easier to acquire. Children respond from a very early age to highly probable content words like *milk*, producing them reliably early in development (Tomasello, 1998). Familiar words contribute to the refinement of phonological categories (Swingley, 2009; Martin, Peperkamp, & Dupoux, 2012) and the acquisition of syntax (Fisher et al., 2010).

The child also notes the frequency of recurring phoneme combinations to pick words out of the speech stream (Saffran, Aslin, & Newport, 1996). Algorithms that attempt to simulate this process, however, sometimes fail to find words or morphemes, and instead under-segment, treating multiword sequences, collocations, and frequent phrases as big words (Feldman et al., 2013; Goldwater et al., 2009). The word segmentation literature sees this result as a failure, but their results raise the interesting possibility that high frequency phrases may be discovered in the same way that common words are, a claim that brings us to the question of phrase frequency effects. If there are common attributes to the representations of these under-segmented phrases and entire words, then "erroneously" treating common phrases as single words may sometimes be useful behavior in language acquisition (Tomasello, 2006) and potentially in its ongoing use by mature language users.

The high frequency phrase advantage

People are sensitive to the frequencies of word sequences, as they are to individual words. One of the first studies to demonstrate phrase frequency effects was Bannard and Matthews (2008). In that study, phrases such as *a drink of milk* and *a drink of tea*, which were matched for semantic class, word frequency, and two-word (bigram) frequency were presented to young children. These phrases differed only in their phrase frequency as measured in a corpus of child-directed speech. *Of milk* is about as common as *of tea*, and *milk* and *tea* are also comparably common in a corpus of British English. However, *a drink of milk* is more common than *a drink of tea*. Recordings of these phrases were played to young children, who were asked to repeat them; they made fewer errors when repeating the more frequent phrases, and were quicker in doing so. These results suggest that children's experience of particular phrases, as well as of words, have a measurable impact on the representations that underlie their developing linguistic abilities.

Adult production seems sensitive to phrase frequency as well. In particular, prosodic measures such as duration reflect the frequencies of multiword combinations as well as the frequencies of the component words. In one study (Arnon & Priva, 2013), frequent 3-word sequences (trigrams) were produced with shorter duration than infrequent trigrams, even when considering word frequencies within those phrases as well. That is, the more frequent a *drink of milk* would have a shorter duration than a *drink of tea*, just as Bannard and Matthews (2008) found in children. Shaoul et al. (2014) used a phrase Cloze completion task to assess implicit knowledge of phrase frequencies. The endings that speakers provided to the incomplete phrases mirrored the phrase frequency distributions that have been observed in corpora. Speech onset times are also sensitive to phrase frequencies. Janssen and Barber (2012) constructed phrases such as *blue car* and *red car* in Spanish, which differed in their phrase frequencies. They asked participants to name pictures that could be described by these phrases. High frequency phrases, but not necessarily phrases containing high frequency words, were initiated more quickly.

Multiword sequences can impact comprehension as well. Smith and Levy (2013) examined whether word and phrase frequencies jointly influence reading times for text, confirming previous findings in reading (Bannard, 2006). They found that there were contributions of word, bigram, and trigram frequencies, such that the more frequent each of these components were, the more quickly those words and word sequences were read. Furthermore, reading times were logarithmic with respect to phrase frequency, an effect that had been robustly demonstrated with words (Howes & Solomon, 1951; Rayner, 1998). This result occurred despite the fact that their statistical model contained no syntactic information, just information about the lexical sequences. Other work on sentence processing suggests that models using information about phrases only can explain reading times just as well as models with syntax (Frank, Bod, & Christiansen, 2011).

Taken together, these results demonstrate that language users represent phrases. At this point, though, there is uncertainty as to the degree to which such representations are holistic. There are many more possible phrases than words. For a vocabulary of N words, there are n^2 bigrams, n^3 trigrams, etc. Thus, a language user often confronts a phrase for the first time, and its meaning will have to be constructed compositionally from its parts (i.e. its words) and from context (e.g. Smith & Osherson, 1984; Medin & Shoben, 1988). Furthermore, there is greater difficulty in estimating the frequencies of phrases, especially those in the lowest frequency ranges (Evert, 2005; Piantadosi, 2014). Given this, efficient encoding of language then might involve representing phrases in a way that makes use of the knowledge of their component words and separately representing only the information that is not contained within the word-level representations (such as frequency of occurrence of the combination).

The contribution of individual words to the fluency of phrase processing is difficult to assess using the previously employed methods. For both words and phrases, higher frequency linguistic events are easier to process and produce than lower frequency events, and the correlation between the frequency of phrases and their component words can make them hard to tease apart. As we noted, a major exception to the general linguistic advantage for high frequency events is apparent in tests of recognition memory, a topic to which we turn now.

A paradox of word frequency

Low frequency words have long been documented to do better on recognition memory tests than high frequency words. Specifically, low frequency words are better identified when they were studied (more hits) and better rejected when they were not studied (fewer false alarms). Crucially, because of the increase in hits *and* the decrease in false alarms to low frequency words, the mirror effect represents a situation that cannot be strictly explained by any one class of words attracting more yes responses than high frequency words, since any such advantage would not play out in opposite advantages for studied and unstudied items.

The word frequency mirror effect is in itself part of a broader set of mirror effect phenomena. In general, words with strange meanings, odd letter combinations, or which occur in only a few contexts in real life, all exhibit the mirror effect (Glanzer & Adams, 1985; Seamon & Murray, 1976; Zechmeister, 1972; Malmberg et al., 2002; Steyvers & Malmberg, 2003). The mirror effect has also been demonstrated for faces varying in typicality (Vokey & Read, 1992) and picture-word pairs that have unusual labels (Bloom, 1971). Malmberg et al. (2002; see also Shiffrin & Steyvers, 1996) attribute the effect to "feature frequency," a conceptualization that suggests that the mirror effect generalizes to any arbitrary distinctive features that are attended to in processing a stimulus, with rare features providing a benefit to memory.

Some accounts of the word-frequency effect in recognition appeal to the impoverished episodic representation of low frequency words. Because people experience low frequency words fewer times than high frequency words, they have more memories of high frequency words. These multiple memories lead generally to the high-frequency advantage in most language processing tasks. But this benefit comes at a cost to memory. We have seen many *cats* but few *panthers* and, consequently, are better able to recover the particular contexts in which we experienced *panther*. It is this recovery of the context of the studied list that is crucial for a recognition decision (Reder et al., 2000). Other accounts have emphasized that the amount of change that our memorial representations undergo is greater for a low frequency word when it is encoded (Benjamin, 2003; Reder et al., 2000). As a result, unstudied low frequency words seem especially novel in comparison to unstudied high frequency words (Benjamin, Bjork, & Hirshman, 1998; Brown, Lewis, & Monk, 1977). Thus, low frequency words benefit from a one-two punch in a recognition memory test. The first effect is that it is easier to recover the studied episode for a low frequency word, leading to more hits. The second effect is that unstudied low frequency words will look especially unfamiliar, leading to fewer false alarms.

If phrases are represented holistically, low frequency phrases should garner more hits and fewer false alarms in a recognition memory test, much like low frequency words. In Experiment 1 we test whether phrase frequency induces a mirror effect in recognition memory in the same way that word frequency does.

Experiment 1a

In this experiment, participants studied 26 adjective-noun phrases that varied in phrase frequency. The studied phrases were sampled from a set of 52. After a 30-minute retention interval, participants saw the complete set of 52 phrases and judged whether the phrase had been previously studied or not.

Method

Participants

Participants were 40 undergraduate students from the University of Illinois who acquired no language before the age of 5 other than English. Participants received course credit for their participation in this experiment.

Materials

52 adjective-noun phrases served as stimuli. These were extracted from the Google 1T n-gram corpus (Brants & Franz, 2006) using word lists for each category extracted from the part-of-speech-tagged British National Corpus (BNC). In these phrases, both the nouns and the adjectives had a restricted frequency range of 19-23.5 (taking the $\log_2$ transform of their frequency in the corpus). The phrases exhibited a relatively broad phrase frequency range ($5.4 < \log_2(\text{phrase frequency}) < 19.7$; see Figure 1.1). The most common phrases (*rheumatoid arthritis*, *alcoholic beverages*) were approximately as common as the least common adjective (*decadent*) and noun (*grasslands*) in our dataset. The stimulus set may be found in Table A of the Appendix.

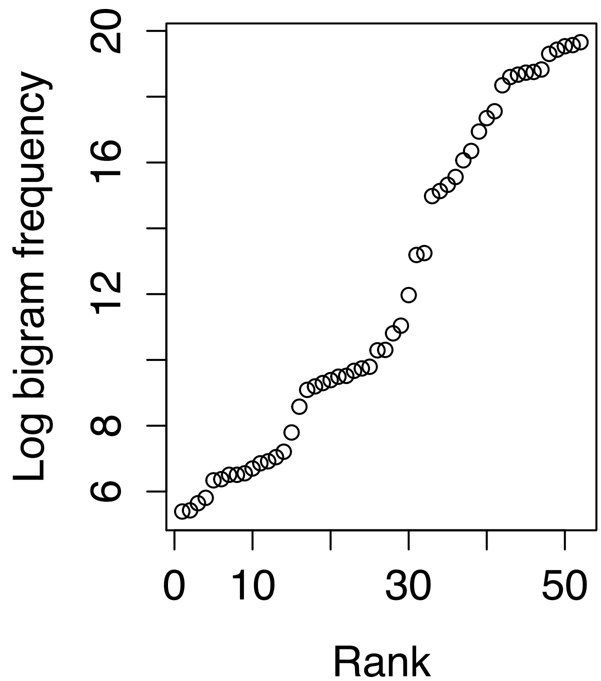

Figure 1.1: Log-frequency rank plot illustrating the uniform distribution of the phrase frequency range for the stimulus set in Experiment 1

The items were selected so that phrase frequency in the materials did not correlate with word frequency (adjectives with phrases, $r = -0.09$, $p = 0.54$; nouns with phrases, $r = 0.17$, $p = 0.23$), which is not normally the case because a common phrase naturally makes its words more common. We also verified the lack of correlation between the two word frequencies ($r = 0.09$, $p = 0.50$). Phrase frequency, noun frequency, and adjective frequency were not correlated with adjective or noun lengths in this stimulus set, and phrase frequency was not correlated with total phrase length. Phrase frequency was neither correlated with orthographic neighborhood density (r = -0.05, p = .73) nor orthographic probability ($r = 0.04$, $p = .76$), which were calculated using the CLEARPOND database (Marian et al., 2012).

Procedure

Each participant received a different set of 26 phrases to study. So that phrase frequency was varied to an even extent in each study set, a randomly seeded sampler selected items using a median split based on the items' phrase frequency, following a method used in prior work (Benjamin, 2003; Tullis & Benjamin, 2012). For each participant, we took random subsets of the top and bottom halves of the phrase frequency range, obtaining 13 phrases from the top, and 13 phrases from the bottom.

Participants were told, "You will be presented with pairs of words that combine to make meaningful phrases that you should memorize. You should try to remember as many of the pairs of words as you can." They were not given further specification about what type of memory test they would complete. The studied phrases were presented in random order. Each phrase was presented at the center of a computer screen for 1 second, with a 1 second inter-stimulus interval. After the study phase, participants put together a puzzle of St. Basil's Cathedral for 30 minutes.

At test, all 52 phrases were presented, again in random order. Each phrase was presented at the center of the screen while participants made a recognition judgment. To make their judgment, they pressed the "p" key if the item was "old" and the "q" key if the item was "new". Participants could take as much time as they wanted to make a response.

Results

The recognition judgments were analyzed using logit mixed effects models. Responses were modeled as a function of whether or not the item being responded to was in the studied list (studied status), the log phrase frequency (phrase frequency), and the interaction of these variables. The random effects included random slopes for the effect of studied status on response bias and intercepts by participants. We also included random intercepts for items. All analyses were completed in the R package `lmer` version 1.1-6 using the optimizer `bobyqa` to prevent non-convergence problems (Bates et al., 2014; Powell, 2009). All coefficients represent changes in log odds of a yes versus a no response as a function of the predictor.

Participants demonstrated the ability to correctly identify studied items ($\beta = 1.63$, $z = 16.92$, $p < .001$). If the predicted greater accuracy for low frequency phrases had occurred, then the interaction coefficient between phrase frequency and studied status would be negative and significantly different from zero. In fact, this interaction was not found and actually was slightly positive ($\beta = 0.08$, $z = 1.28$, $p = .20$). What we saw instead was only a bias for participants to say that they had studied high frequency phrases, regardless of whether the phrase had been studied or not ($\beta = 0.39$, $z = 4.55$, $p < .001$). We illustrate the bias effect that phrase frequency has on hits and false alarms in Figure 1.2. The full model is reported in Table 1.1.

Table 1.1
Summary of Experiment 1a fixed effects

Predictor	Parameter estimates		Wald's test	
	Log-odds	S.E.	Z	p
(Intercept)	-0.46	0.12	-3.95	<. 001
Old or New Status	1.63	0.10	16.92	< .001
Phrase frequency (bias)	0.39	0.09	4.55	< .001
Phrase frequency by Old-New Status	0.08	0.06	1.28	.20

Note: Significance obtained at $p < .05$.

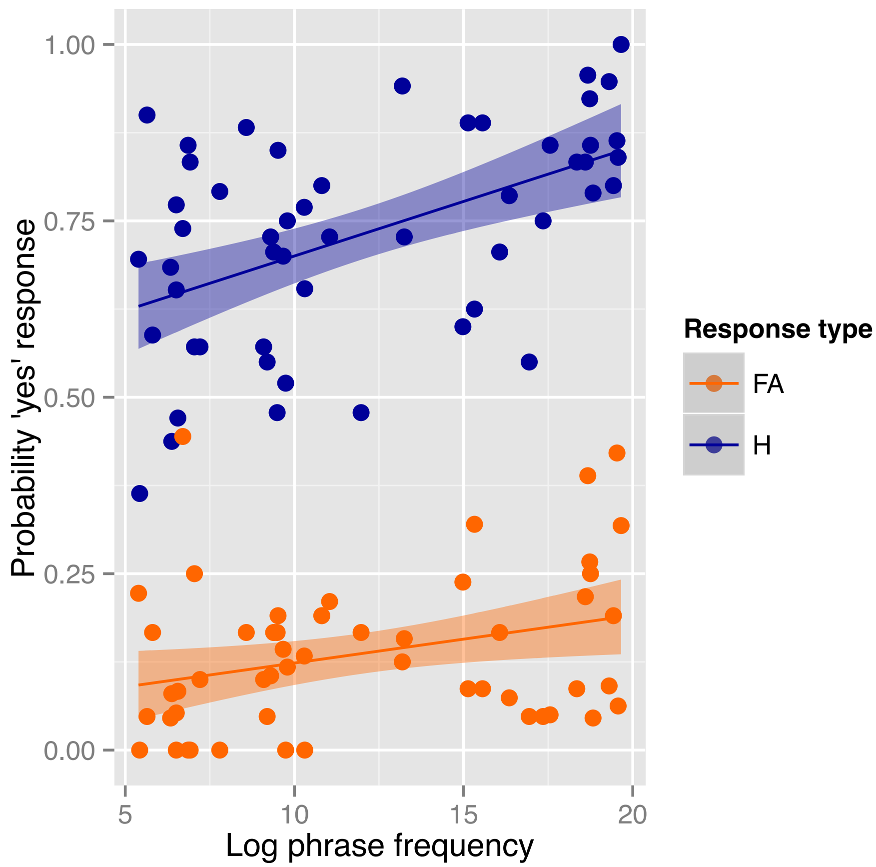

Figure 1.2: Hit rates and false alarm rates to phrases for Experiment 1a as a function of phrase frequency, collapsed across participant variance. The shaded areas correspond to one standard error around the regression line. Participants make more hits and false alarms to high frequency phrases.

Discussion of Experiment 1a

The lack of a phrase frequency mirror effect suggests that the effect of phrase frequency on participants' representation of their language is different from the effect of word frequency, at least with respect to recognition memory. This effect is surprising, because many stimuli benefit from some kind of "unusualness" in recognition memory tasks

(Glanzer & Adams, 1985; Malmberg et al., 2002; Steyvers & Malmberg, 2003; Vokey & Read, 1992; Seamon & Murray, 1976). We note, though, that the bias to respond "yes" as phrase frequency increases does suggest that frequency influences performance. Therefore, speakers do somehow encode information about the frequency of the word combinations.

Experiment 1b

The unexpected results of Experiment 1a motivated an attempted replication. In Experiment 1b, we looked again for a phrase frequency mirror effect. We also sought to rule out the possibility that the bias toward saying "yes" to high frequency phrases was due to the simultaneous presentation of the two words of the phrases at study and test. Presenting the words in sequence could encourage the separate processing of the individual words and possibly nullify the phrase frequency bias effect of Experiment 1a. Because of these concerns, we repeated Experiment 1a in all respects, except that the two words of the phrases were presented in sequence during study and at test.

Method

Participants

Participants were 40 undergraduate students from the University of Illinois who acquired no language before the age of 5 other than English. Participants received $8 for their participation in this experiment.

Materials and Procedure

The only difference between Experiment 1a and Experiment 1b was the manner in which the phrases were presented at study and at test. In Experiment 1b, we presented phrases word by word, instead of simultaneously. At the beginning of every study trial, we presented the adjective at the center of the screen for 450ms, followed by a 50ms blank screen before presenting the noun at the center of the screen for 450ms. Words within phrases never appeared together. There was a 1 second inter-trial interval before the presentation of the next phrase. To the extent that participants chose to encode the pairs of words as phrases, it would likely be due to the longer intertrial interval between phrases than the interstimulus interval between words in a phrase. After study, participants again put together a puzzle for 30 minutes.

Presentation of the phrases at test followed a similar design. Participants were asked to respond as to whether the phrases presented were ones they had studied or not. The rate of presentation of the words within the phrases was the same as at study, with the adjective and noun on the screen at separate times. In addition, judgments were solicited only after both words had been presented and removed from the screen. Only the cues as to what response to make (“p” for “old” and “q” for “new”) were on the screen during the response. Participants were told to judge whether they had studied the entire phrase. They were allowed to take as much time as they needed to make a response.

Results

Analysis followed as in Experiment 1a. We again found no low frequency advantage, as evidenced by the lack of interaction between phrase frequency and studied status (β =0.05, z = 0.65, p = .26). Also as before, we found that participants were biased toward saying that they had studied high frequency phrases, though this effect was somewhat weaker in this experiment than in Experiment 1a (β =0.27, z = 2.31, p < .05). We illustrate the phrase bias in Figure 1.3. The full model is reported in Table 1.2.

Table 1.2
Summary of Experiment 1b fixed effects

Predictor	Parameter estimates		Wald's test	
	Log-odds	S.E.	Z	p_z
(Intercept)	0.09	0.18	0.62	.73
Old or New Status	1.77	0.13	14.14	< .001
Phrase frequency (bias)	0.27	0.12	2.31	< .05
Phrase frequency by Old-New Status	0.05	0.08	0.65	.26

Note: Significance obtained at $p < .05$.

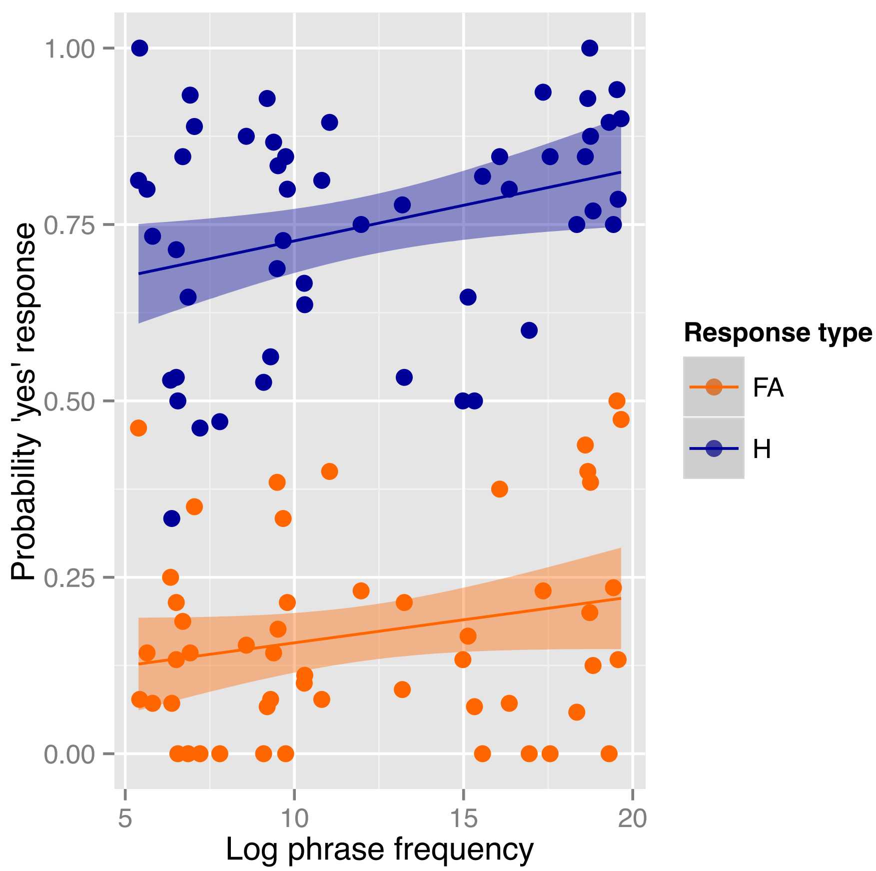

Figure 1.3: Hit rates and false alarm rates to word-by-word presented phrases for Experiment 1b as a function of phrase frequency, collapsed across participant variance. The shaded areas correspond to one standard error around the regression line. As in Experiment 1a, participants show more hits and false alarms to high frequency phrases.

Discussion of Experiment 1b

In Experiment 1b, we replicated the findings of Experiment 1a. There is evidence in both experiments that participants use phrase frequency to make their judgments about whether a phrase was studied or not (evidenced in a bias to say "yes" to more common phrases), but phrase frequency does not appear to impact people's ability to discriminate studied from unstudied phrases. The fact that the results of 1a replicated even when the words are presented individually suggested that the phrases are processed as units.

Interim Discussion

Given the results of Experiments 1a and 1b, we propose the following model of phrase frequency effects, which is outlined in Figure 1.4. We base our model on episodic accounts of the word frequency effect in recognition memory, most specifically Reder et al., (2000), as other models of frequency effects (e.g. Baayen et al., 2013) have not been developed for recognition memory. Specifically, we propose that each experience with a multiword sequence leaves a trace (with each episode represented by a star in the figure). For example, consider the phrase *psychic nephew*. Each experience with this particular phrase results in another episodic token. This includes the experience of studying *psychic nephew* in an experimental list (represented by the red star).

The Reder et al. model accounts for the higher hit rates for low frequency items and higher false alarm rates for high frequency items using two mechanisms. Because an individual episodic memory has fewer competitors for low frequency items, the study episode for that item is more likely to be chosen at test. Higher false alarm rates for unstudied high frequency words arise because the baseline activation of the word is higher. In sum, when high frequency words are studied, the individual episode is more difficult to retrieve among competitor episodes, but when the high frequency word is new, the baseline activation or familiarity of that item is already high, leading to a bias to say yes to that item even when it was unstudied. Critically for phrase memory, these episodic tokens can be retrieved from memory not just from a cue that matches the entire phrase, but from a cue that matches part of it, such as the noun. So, the tokens can be thought of sets of as features that represent the experience of the phrase, with a featural cue having the capacity to retrieve an entire episode. Crucially, words act as features. Thus, this model takes a compositional, episodic approach. Relevant phrasal episodes are retrieved because the episodes contain their words, and the influence of frequency is attributed to the multiplicity of episodes.

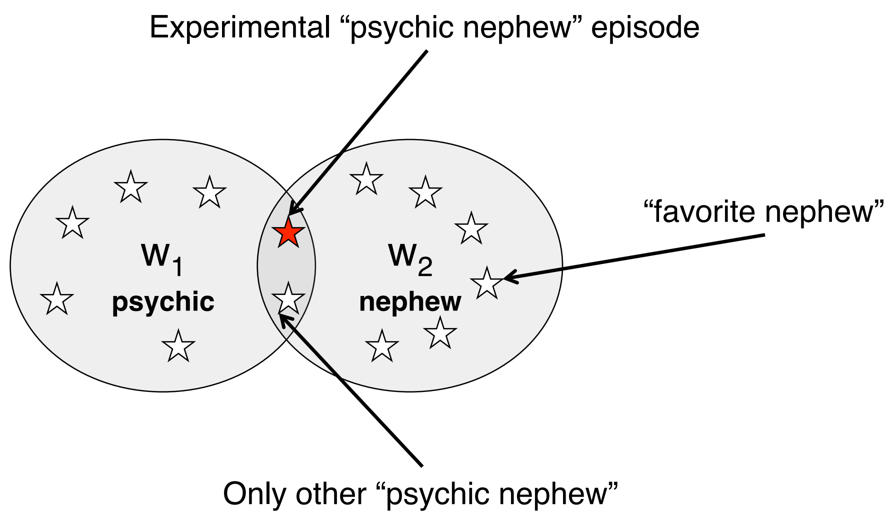

Figure 1.4: Schematic for the representation of the episodic memories involved in the multiword phrase *psychic nephew.* Each word within the phrase is represented as a circle containing episodic memories (stars). So, the word *nephew* might have many other memories, such as *favorite nephew*. When a phrase is processed at study, the red star is

placed at the intersection of the two words. The process of recognition at test requires retrieving the red star that was experienced during study. At test, all memories associated with the words compete for retrieval, so words with many more memories make that phrase more difficult to locate. Because phrases are generally impoverished (the intersection between two words is often very unpopulated), phrases have very few competitors from the same phrase, so word frequency becomes more important.

Why is there no benefit for low frequency phrases? In a recognition test, each word of a test phrase serves as a retrieval cue. So, when *psychic nephew* is presented at test, it has the potential to retrieve all episodes with *psychic* and all episodes with *nephew*, as indicated by circles in Figure 1.4. We assume, consistent with prior work (e.g. Smith & Osherson, 1984), that nouns contribute more to the meaning of an adjective-noun phrase, so episodes that overlap in just the noun will be more retrievable than those that share only the adjective. As in the model of Reder et al. (2000), recognition judgments are determined in part by whether or not the critical episode (the red star representing that the phrase that was studied in the experiment) has been retrieved. Finding that episode is more difficult when many other episodes are active. Because words are far more frequent than phrases on average, the main determinant of the number of interfering episodes will be the frequencies of the *words* within a phrase, particularly the noun, and not the phrase itself. Because there are relatively few memories of the whole phrase, they contribute few interfering episodic tokens. In fact, for many phrases, the number of possible episodic tokens is possibly zero (e.g. *psychic nephew*), so only word-level information would be available for use during search for the critical episode.

Why is there a bias to say “yes” for high frequency phrases? Reder et al. (2000)’s word-frequency model assumes that there exist abstract representations of word types in addition to episodic memories. When a word is more frequent, this representation is stronger and contributes to a feeling of familiarity, and thus to a bias to say yes. We can borrow the same account for phrase frequency. This requires that there be an abstract holistic representation of the phrase (sensitive to frequency) in addition to the hypothesized phrasal episodes. The representation of strength does not necessarily have to be a property of a phrasal "node." For example, it could be an association strength between, say, the adjective and the noun, such as might be acquired from a model that learns through prediction error about subsequent words, given previous ones. Alternately, we can dispense with abstractions and hypothesize that somehow, the participant is able to discern phrasal familiarity from the set of retrieved episodes that match on both words (e.g. the number of episodes that contain both *psychic* and *nephews*). Our data do not distinguish between this episodic and abstractionist account of the bias to say “yes”.

The model we outline generates a specific prediction from its assumption that phrasal episodes have a compositional nature: The frequency of the words within a phrase should affect the amount of interference that is generated at test, such that phrases containing low frequency words should be better recognized than phrases containing high frequency words, leading to a word frequency mirror effect. We specifically expect to see a contribution of noun frequency to phrase memory because of the greater contribution of the noun to phrase meaning. A phrase with a frequent noun should tend to attract fewer hits and more false alarms than a comparable phrase with a less common noun. In the next section we run a combined analysis of Experiments 1a and 1b to look for preliminary evidence of a noun frequency mirror effect.

Cross-experiment analysis

Norming study. Phrases, like words, have conceptual properties associated with them that may enhance or obscure memory for those phrases. In particular, phrases differ from monomorphemic words by having meanings that can be composed, or which are idiomatic (e.g. *red house* versus *red herring*). However, like words, phrases may be familiar concepts or not. It is necessary to ask, therefore, whether the effects of phrase frequency that we saw in Experiment 1a and 1b might be in part due to the relationship between these factors and phrase frequency. To account for these factors, we conducted an additional norming study with 50 participants from the University of Illinois course credit subject pool. Each participant rated each of the 52 phrases for concreteness (e.g. "This phrase denotes a real-world entity"; Paivio et al., 1968), imageability (e.g. "I can easily picture what this phrase describes."), and compositionality (e.g. "Are alcoholic beverages beverages that are alcoholic?"; Szabo, 2013) on a three-point scale ("Not at all", "Somewhat", and "Definitely"). We then averaged over all 50 participants for each of the 52 items to obtain concreteness, imageability, and compositionality scores to use as control variables.

We first constructed a null model using the results of the norms to predict memory performance, and then introduced our key predictors: phrase frequency, noun frequency, and adjective frequency. Concreteness and imageability were highly correlated ($r = .93$), while compositionality was less strongly correlated with concreteness ($r = .62$) and imageability ($r = .69$). Due to these correlations, we only added imageability and compositionality, and their potential interactions with studied status, to the null model. Because this analysis uses the data from both Experiments 1a and 1b, we added *Experiment* as a fixed effect. *Experiment* did not significantly interact with any terms in the model, so we did not retain the higher order interactions, and only include Experiment as an additive term in the null model. We then conducted a stepwise additive model building procedure to test for differential effects of phrase frequency, noun frequency, and adjective frequency on hits and false alarms. Random effects terms with near-perfect correlations were removed (Baayen et al., 2008).

First, we introduced phrase frequency and its possible interaction with studied status. This significantly improved model fit over the null model (χ^2 (7) = 62.48, $p < .001$). We then found that the addition of a noun frequency main effect term as well as the interaction of noun frequency with studied status again improved model fit (χ^2 (11) = 19.89, $p < .05$). Finally, we asked whether adjective frequency contributed anything to model fit. The adjective terms did not significantly improve the likelihood of the model (χ^2 (2) = 2.56, $p = .28$), so adjective frequency and its interaction with studied status were not included in the final model. The final model is presented in Table 1.3.

Altogether, the results suggest that phrase and noun frequency contribute to recognition memory judgments. First, participants are more likely to say "yes" to a higher frequency phrase than a lower frequency phrase, regardless of whether the item was actually studied or not ($\beta = 0.36$, Wald $Z = 3.77$, $p < .001$). This result shows that the phrase-frequency bias effect identified in each of the two experiments is robust when phrasal differences in compositionality and imageability are taken into account.

Critically, phrases containing uncommon nouns show a benefit to recognition memory, as evidenced by a noun frequency mirror effect (β = -0.41, Wald Z = -3.98, p < .001). This was exactly what we predicted from our model. This suggests that memory for phrases depends at least in part on the distinctiveness of the component parts, specifically the nouns, which are central to the meaning of the phrase and have been implicated in prior research as an "anchor" in memory (Yuille, Paivio, & Lambert, 1969; Richardson, 1978; Morris & Reid, 1972; Lockhart, 1969; Mata, Percy, & Sherman, 2013). We note one additional finding from the final model: As has been reported previously in the literature (Paivio, 1971), increasing imageability led to greater phrase discriminability (β = 0.18, Wald Z = 2.62, p < .01).

Table 1.3
Summary of Experiment 1a and 1b combined analysis

Predictor	Parameter estimates		Wald's test	
	Log-odds	S.E.	Z	p_z
Intercept	-0.26	0.11	-2.31	< .05
Studied status	1.71	0.08	22.11	< .001
Phrase frequency	0.36	0.09	3.77	< .001
Experiment	0.66	0.18	3.75	< .001
Noun frequency	0.05	0.10	0.52	.30
Compositionality	-0.08	0.12	-0.66	.25
Imageability	-0.06	0.12	-0.51	.30
Phrase frequency * Studied status	0.08	0.05	1.52	.06
Noun frequency * Studied status	-0.23	0.06	-3.98	< .001
Imageability * Studied status	0.18	0.07	2.62	< .01
Compositionality * Studied status	0.01	0.07	0.14	.55

Note: Significance obtained at *p* < .05.

The presence of a noun frequency mirror effect provides preliminary support for our account. It generally suggests that knowledge of words contributes to the processing of phrases, and thus that phrasal representations are not entirely holistic.

Experiment 2

The phrases used in Experiment 1 were taken from the Google n-gram corpus as described. While this tells us that they occurred on the Internet with some frequency, many of the infrequent phrases (e.g. “chrome throttle" or "psychic nephew”) would not be encountered frequently in daily life, and consequently we cannot be sure that they are meaningful to participants. This may put them at an encoding disadvantage, as has been seen in recognition memory for pseudowords (e.g. Diana & Reder, 2006). We therefore tested whether our key effects hold for another set of phrases where even the "low frequency" phrases are likely to be familiar and meaningful to participants.

To that end, we developed an additional stimulus set from the spoken portion of the Corpus of Contemporary American English (COCA; Davies, 2008), which consists primarily of publically broadcast material from the news, on talk shows, etc. These phrases therefore represent more easily recognizable phrases. We gathered a total set of 112 phrases (56 in a high frequency phrase list and 56 in a low phrase frequency list) meeting several criteria, which we discuss below.

All the phrases we gathered from COCA were compositional (nonidiomatic) adjective-noun phrases varying in their frequency of occurring in the subset of the database containing spoken English. We calculated the spoken frequencies of these phrases from the years 2009 to 2012, which represents more a recent and ecologically valid sample of the language the typical freshman undergraduate would experience while watching the news from the beginning of middle school through the most recent collection of data in COCA. Noun and adjective length did not significantly correlate with phrase frequency (Pearson's $r = -.11$, $p = .28$ and $r = -0.14$, $p = .16$).

Nouns and adjectives were deliberately selected to be higher in frequency than in Experiment 1 in order to increase the chances that the participants actually knew all of the words within the phrase, with the least common adjective and noun occurring 200 times more often than the least common phrase. Frequencies for the adjectives and nouns were restricted to the same range, from 1031 to 4021 and from 1026 to 4037, respectively out of the entire corpus from 2009 to 2012. As such, all nouns and adjectives were within a single power of 2 in COCA frequency. The lowest frequency phrases were "poor credit", "southern food", "fantastic panel", and "nice hair". The highest frequency phrases were "foreign language", "presidential candidate", "middle class", and "grand jury". Log_2 frequencies of the counts ranged from 2.32 to 9.57. These phrases are listed in Appendix B.

Procedure

The procedure was the same as in Experiment 1a.

Results

The analysis proceeded as in Experiment 1. We replicated the key results of that experiment[1]. There was a main effect of studied status, suggesting that participants were highly accurate (B = 2.60, z = 9.15, p < .001. There was no interaction between studied status and phrase frequency (B = -0.14, z = -1.07, p = .28), indicating that there was no frequency-related mirror effect. Crucially, there was a main effect of log phrase

1 The contribution of the noun to phrase memory with the materials from the COCA corpus was unsurprisingly quite small, as noun (and adjective) frequencies were quite restricted. There was neither a main effect of noun frequency (B = -0.02, z = -0.23, p = .82) on recognition responses, nor an interaction between noun frequency and studied status (B = 0.05, z = 0.47, p = .63).

frequency on whether participants were likely to call a phrase old or new ($B = 0.19$, $z = 2.09$, $p < .05$). When a phrase was high frequency (e.g. "foreign language") participants were more likely to say it was studied than a low frequency phrase (e.g. "angry crowd") regardless of whether the phrase had been studied or not. These results are summarized below in Table 1.4 and plotted below in Figure 1.5.

Table 1.4
Summary of Experiment 2 fixed effects

Predictor	Parameter estimates		Wald's test	
	Log-odds	S.E.	Z	p_z
(Intercept)	-2.06	0.25	-8.35	< .001
Old or New Status	2.60	0.28	9.15	< .001
Phrase frequency (bias)	0.19	0.09	2.09	< .05
Phrase frequency by Old-New Status	-0.14	0.13	-1.10	.28
Noun frequency (bias)	-0.02	0.09	-0.23	.64
Noun frequency by Old-New Status	0.05	0.11	0.47	.82

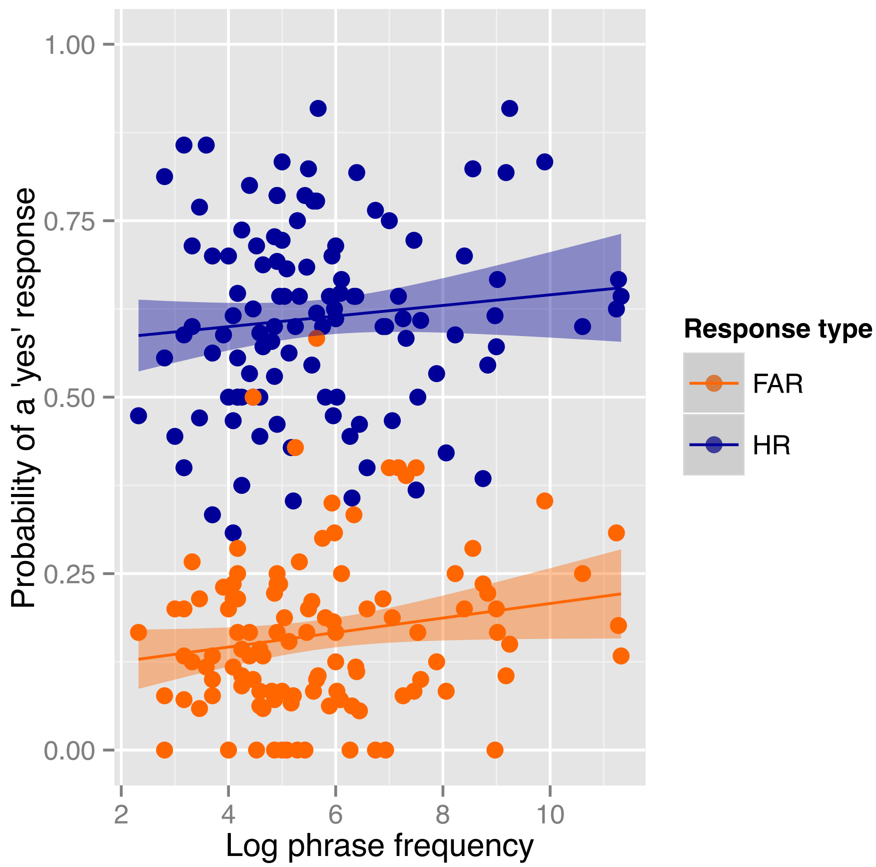

Figure 1.5: Hit rates and false alarm rates to word-by-word presented phrases for Experiment 2 as a function of phrase frequency in COCA, collapsed across participant variance. The shaded areas correspond to one standard error around the regression line. As in Experiment 1a and 1b, participants show more hits and false alarms to high frequency phrases.

Discussion of Experiment 2

The results of this experiment demonstrate that the bias to endorse high-frequency phrases as having been studied is not an artifact of the stimuli from Experiment 1, some of which may have been nonsensical to some subjects. We see the same pattern of results in this experiment as we do in Experiment 1: high frequency phrases are more likely to garner "yes" responses regardless of whether the phrase was studied or not. Furthermore, this experiment, like Experiment 1, failed to show the frequency-based mirror effect that is typically observed in recognition memory experiments for single words.

Experiment 3a

The results from Experiments 1a and 1b suggest that noun frequency controls our ability to discriminate studied from unstudied phrases. On the other hand, phrase frequency seems to have some effect on the impression of familiarity for the phrase without affecting accuracy, as seen in both Experiment 1 and 2. The importance of the noun for phrase memory is not without precedent: some work suggests that letter frequencies can lead to the mirror effect, with words with uncommon letters garnering more hits and fewer false alarms (Malmberg et al., 2002). In our case, the uncommonness of the noun contributes to the discriminability of a phrase in recognition memory. The next two experiments sought to confirm this finding. In Experiment 3a we determined the strength of the relationship between word frequency and recognition with single words (nouns), and then in Experiment 3b embedded those same words in phrases with the goal of providing a definitive test of our prediction. Because we did not explicitly manipulate phrase frequency in these experiments, the results of these two experiments can only speak to the role of the noun in phrase memory as a test of our account.

Method

Participants

Participants were 30 undergraduate students from the University of Illinois who acquired no language before the age of 5 other than English. Participants received $8 for their participation in this experiment.

Materials

Eighty-eight nouns from a set of ninety-six nouns used by Balota et al. (2002) served as the stimuli, which had been controlled for concreteness/imageability and word length. The full set of nouns was not used because there are additional constraints based on phrase construction that will be clarified when we introduce Experiment 3b. Words in the dataset spanned a continuous frequency range of 16.9-28.4 in $\log_2$ and included, for example, *tree*, *wizard*, and *anvil*. All nouns were concrete with the exception of *nation*. The materials for Experiment 3a are found in the "noun" column of Appendix C.

Procedure

The 88 nouns were repartitioned based on their frequencies in the Google corpus into "high" and "low" frequency categories based on a median split. This split resulted in some items from the Balota et al. (2002) materials, which had been assigned to "low" and "high" frequency categories, switching frequency categories. A random sample of each half of the high and half of the low frequency nouns comprised the study materials, for a total of 44 study items.

As in Experiment 1a, each noun was presented for 1 second, followed by a 1 second inter-stimulus interval. Due to the greater number of items at study and at test than in Experiment 1, there was no retention interval prior to starting the test. Participants then completed a yes-no recognition test where the nouns were presented and remained on the screen until participants responded. Participants could take as much time as needed to make a response.

Results

We again modeled participant responses to each item as a function of whether the noun was studied or not, noun frequency, and the interaction of those two terms. The most important result was that there was a strong noun frequency mirror effect, such that low frequency nouns received significantly more hits and fewer false alarms (β = -0.48, z = -6.93, p < .001). Unlike the two previous experiments, participants did not exhibit a bias to respond positively (or negatively) as a function of the frequency of the test item (β = -0.06, z = -0.63, p = .27). These results are summarized in Table 1.5. A visual inspection reveals a strong relationship between hit and false alarm rates and noun frequency, which we include in Figure 1.6.

Table 1.5
Summary of Experiment 3a fixed effects

Predictor	Parameter estimates		Wald's test	
	Log-odds	S.E.	Z	p_z
(Intercept)	-0.42	0.15	-2.77	< .001
Old or New Status	1.86	0.14	13.59	< .001
Noun frequency (bias)	-0.06	0.10	-0.63	.27
Noun frequency by Old-New Status	-0.48	0.07	-6.93	< .001

Note: Significance obtained at *p* < .05.

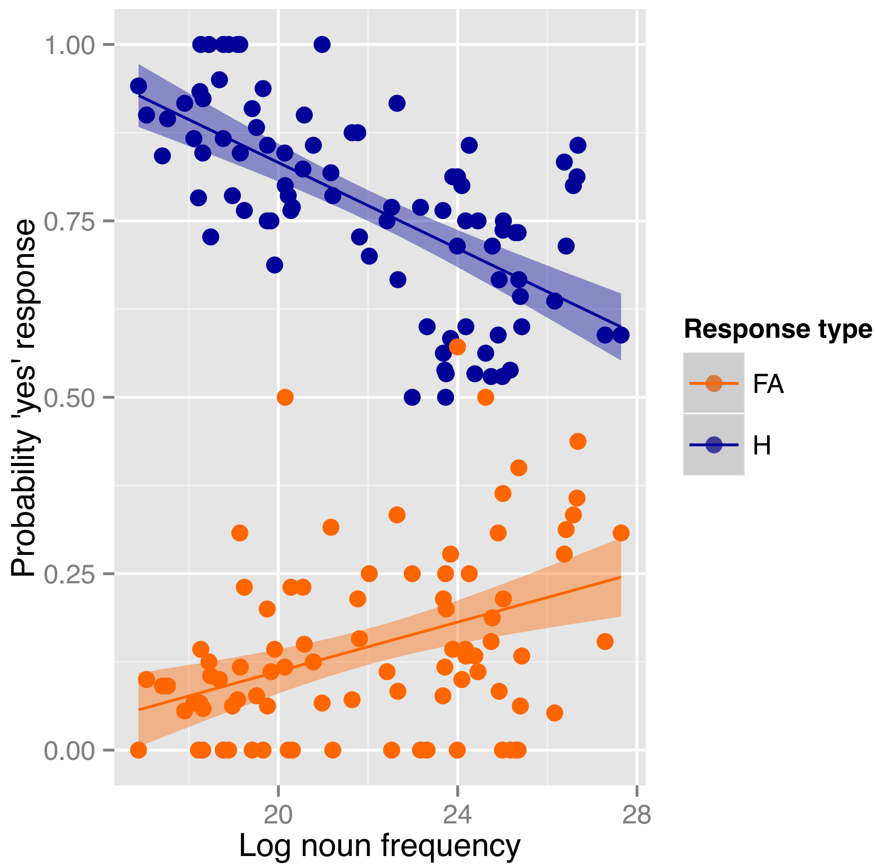

Figure 1.6: Hit rates and false alarm rates to nouns for Experiment 3a as a function of noun frequency, collapsed across participant variance. The shaded areas correspond to one standard error around the regression line. Participants make more hits and fewer false alarms to low frequency words.

Discussion of Experiment 3a

The results of this experiment demonstrate that the word frequency mirror effect for our items is robust. Given this, the nouns were then incorporated into adjective-noun phrases to evaluate the degree to which this relationship holds when those phrases do not vary in adjectival or phrase frequency. Specifically, we looked for an effect of noun frequency even when the study of those nouns is incorporated in phrases.

Experiment 3b

Method

Participants

Participants were 30 undergraduate students from the University of Illinois who acquired no language before the age of 5 other than English. Participants received $8 for their participation in this experiment.

Materials

The items used in this experiment were adjective-nouns phrases containing the nouns from Experiment 3a. We created these phrases using a corpus of part-of-speech tagged adjective-noun phrases within the Google 1T n-gram corpus (Brants & Franz, 2006). The adjectives and nouns were identified using part of speech labels available from the BNC. The process of excluding nouns that did not occur in our subset of the Google corpus limited the set of nouns used to Experiment 3a to 88. We used 88 adjective-noun phrases in Experiment 3b that contained the nouns tested in Experiment 3a.

We chose the adjectives in these phrases from a very narrow frequency distribution (within a unit of $\log_2$ frequency). Moreover, when these adjectives were combined with the nouns, the resulting phrases also had a very narrow frequency distribution (within a factor of 2). There were no significant correlations between any of adjective, noun, or phrase frequencies, and the means and ranges of all frequencies are almost identical. This was equally true when we divided the nouns into *high* and *low* frequency halves. We present these summary statistics in Table 1.6. Such resulting phrases included *handsome wizard* (containing a low frequency noun) and *premature tree* (containing a high frequency noun). These are available in Appendix C.

Table 1.6
Ranges of (log_2) frequencies by noun frequency category in Experiment 2

Noun frequency	Mean adjective frequency	Adjective frequency range	Mean phrase frequency	Phrase frequency range
Low	22.28	21.5-23.5	7.38	6.75-8
High	22.33		7.4	

Procedure

Participants studied and were tested on adjective-noun phrases containing the nouns from Experiment 3a. Materials were sampled for each participant in the same way as in Experiment 3a. This experiment followed the same study and test procedures as in Experiment 1a, so participants studied and then were tested on phrases with both the adjective and noun presented simultaneously. As in the prior experiment, there was no retention interval.

Results

The analysis of this experiment was the same as in Experiment 3a. Because we only manipulated noun frequency, while holding the other factors constant, the only frequency factor that was considered was noun frequency. Crucially, and as predicted by our account, there was a noun frequency mirror effect, such that phrases containing low frequency nouns got more hits and fewer false alarms than phrases containing high frequency nouns (β = -0.25, z = -2.52, p < .05), although this effect was considerably more modest than in Experiment 3a, in which the nouns were presented and tested alone. Also, as in the previous experiment, they showed no frequency-related response bias; that is, they were not significantly more likely to say that they had seen phrases containing high frequency nouns (e.g. *premature tree*) than low frequency nouns (e.g. *handsome wizard*; $\beta = 0.07$, $z = 0.69$, $p = .25$). We summarize these results in Table 1.7 and the data are pictured in Figure 1.7.

Table 1.7
Summary of Experiment 3b fixed effects

Predictor	Parameter estimates		Wald's test	
	Log-odds	S.E.	Z	p_z
(Intercept)	-0.83	0.17	-4.98	< .001
Old or New Status	2.10	0.17	12.17	< .001
Noun frequency (bias)	0.07	0.10	0.69	.25
Noun frequency by Old-New Status	-0.25	0.10	-2.52	< .05

Note: Significance obtained at $p < .05$.

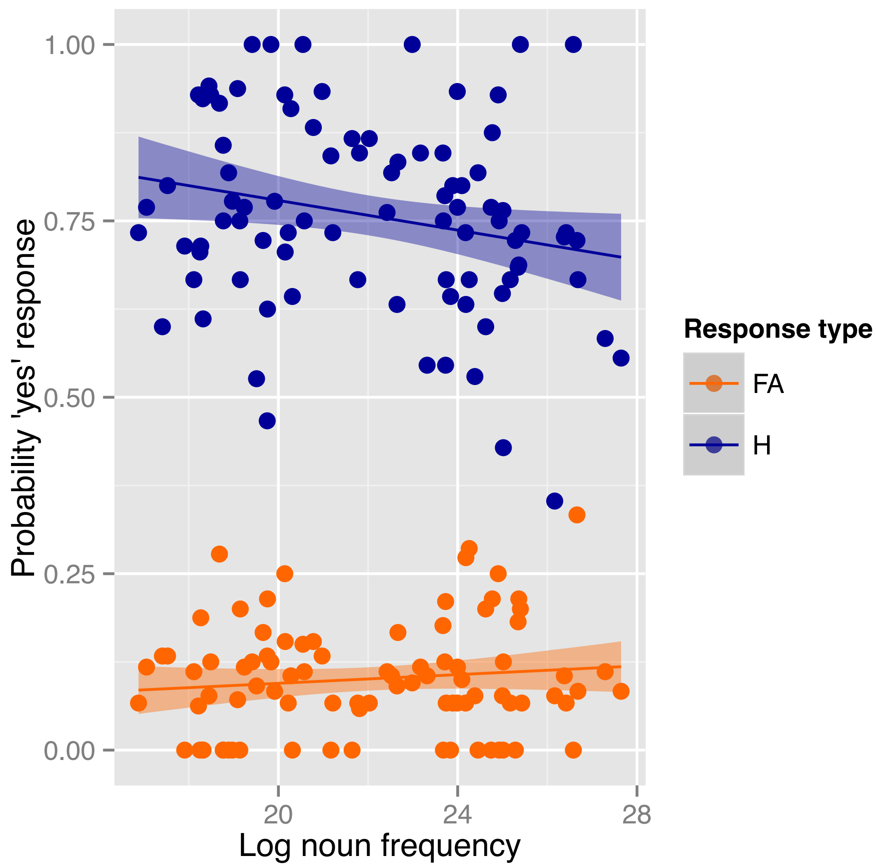

Figure 1.7: Hit rates and false alarm rates to phrases for 3b as a function of noun frequency, collapsed across participant variance. The shaded areas correspond to one standard error around the regression line. Participants make more hits and fewer false alarms to phrases containing low frequency words.

As a final test, we assessed whether the nouns' memorability in Experiment 3a was a predictor of performance on phrases containing those nouns in Experiment 3b. Because of the nature of our model, we predict that phrases containing more memorable nouns should be better recognized. We assessed this using a simple linear regression analysis relating the discriminability (d'; Verde, MacMillan, & Rotello, 2006) of the phrases to the discriminability of the nouns. We found a reliable relationship between noun memorability and phrase memorability, with phrases containing more memorable nouns being better recognized (Pearson's r = -0.24, SE = 0.10, $p < .05$), summarized in Table 1.8 and Figure 1.8.

Table 1.8
Summary of analysis relating phrase discriminability to noun discriminability

Predictor	Parameter estimates			
	Pearson's r	S.E.	*t*	*p*
(Intercept)	-0.82	0.16	-5.06	< .001
Noun discriminability	-0.24	0.10	-2.48	< .05

Note: Significance obtained at $p < .05$.

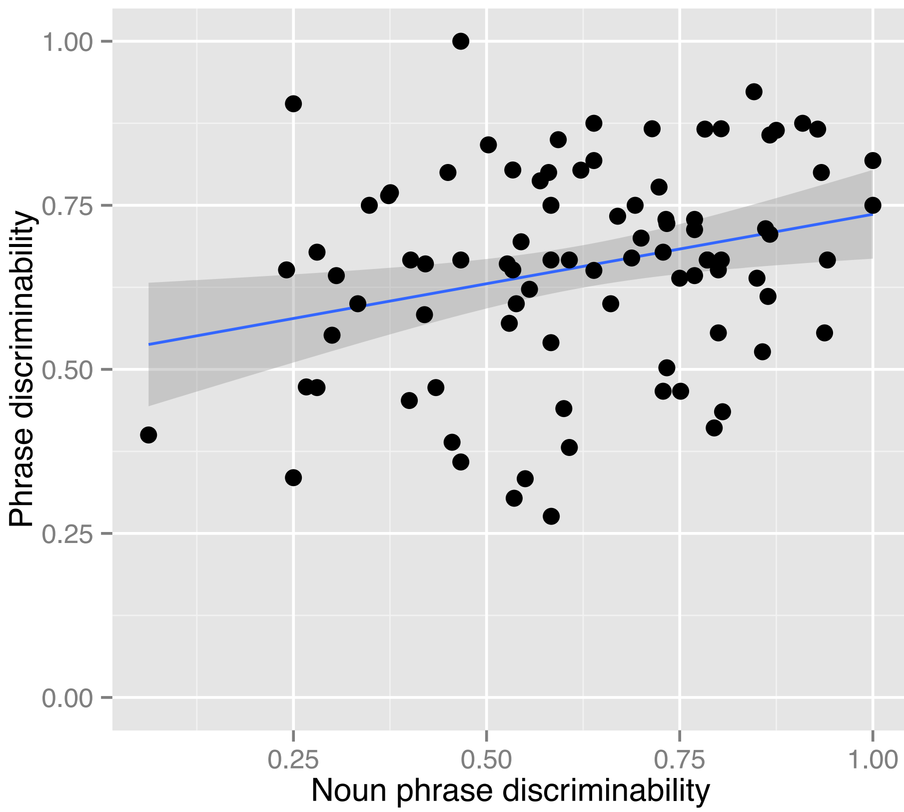

Figure 1.8: The discriminability of a phrase as a function of the discriminability of the noun within the phrase. Phrases containing nouns that are more memorable (higher discriminability) are discriminated better as well, as predicted by the model.

Discussion of Experiment 3b

This experiment suggests that, as with letters within words (Malmberg et al., 2002; Zechmeister, 1972), words within phrases can provide a cue to memory about whether that phrase was studied or not. Furthermore, the results confirm the effects seen in the joint analysis of Experiments 1a and 1b, where we found a small noun frequency mirror effect. This result was a key prediction of the theoretical position outlined earlier.

CHAPTER 2: RECALL

In many linguistic tasks, phrase frequency effects mirror word frequency effects, in which common words (e.g. woman) and phrases (e.g. alcoholic beverage) are easier to acquire, understand and produce than uncommon words and phrases (Janssen & Barber, 2012; Arnon & Snider, 2010; Arnon & Cohen Priva, 2013; Arnon & Cohen Priva, 2014; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Morgan & Levy, 2016; Bybee, 2006; Bannard & Matthews, 2008). Such effects demonstrate that the language processing system pays attention to multiword linguistic units. Frequency effects for individual words have typically been accounted for by either positing a lexical entry that keeps track of something like the count of times a person has encountered a linguistic category or individual memories (exemplars, episodes, or instances) for each of those experiences. Because phrases include a temporal or grammatical relationship between multiple words, it is less clear how phrases might be represented in long-term memory. The present study addresses this question.

One way to explain phrasal frequency effects and phrase representation in general is to propose the existence of a lexically-specific but usage-event-independent representation of the phrase, such as a “node” (e.g. MacKay, 1982) or “superlemma” (e.g. Sprenger, Levelt & Kempen, 2006) that contains information about its category (e.g. noun phrase, for an adjective-noun combination) and connects to representations of its component words (e.g. Copestake et al., 2002). The frequency of a phrase could be stored with this lexical entry, or it could arise from the number of stored episodes that contain or point to it. Alternatively, phrases could lack explicit discrete representations entirely, in line with theories and computational models that encode all words and phrases implicitly in network weights (Seidenberg & McClelland, 1989; Baayen, Hendrix, & Ramscar, 2013; Baayen, Milin, Đurdević, Hendrix, & Marelli, 2011).

Some recent work has looked into whether phrase frequency effects are a product of the mass action of episodic representations of events in which the phrase is experienced. Chapter 1 tested whether people showed the same sensitivity to phrase frequency in recognition memory as they are known to have for words. In single-word recognition memory experiments, words that a participant has rarely experienced over the course of their life (low frequency words) have fewer episodic memories in long-term memory, and yet are more accurately discriminated from lures than high frequency words are (Reder et al., 2000; Hintzman, 1988; Glanzer & Adams, 1985). This paradoxical effect of word frequency can be explained by noting that to judge a test word as "old" in a recognition task, the participant may retrieve the episode in which the word was studied. When that word is low frequency, there are fewer other episodes of it to hinder the search for the crucial experimental episode. In Chapter 1 we reasoned that, if adjective-noun phrases have their own episodic memories that contribute to memory in the same manner, then low frequency phrases like *psychic nephew* should also be more accurately recognized than high frequency phrases like *alcoholic beverages*. Surprisingly, they found that high and low frequency phrases were recognized equally well, but that recognition memory improved when the noun in a phrase was uncommon (e.g. *infarction* in *myocardial infarction*). They concluded that recognition judgments for phrases are more influenced by the number of episodes containing particular words within the phrase, as opposed to the entire phrase. This is likely because individual words are necessarily much more common than phrases. Thus, the many episodes sharing a word with a test phrase are more potent sources of interference in the recognition process than the few episodes containing the entire phrase.

This finding from Jacobs et al (2016) provides evidence that phrasal processing is at least partially compositional, in that judgments about *psychic nephew* are influenced by memories of events of psychic things that are not nephews and nephews that are not psychic. However, the study also found that participants tended to say they had studied the more common phrases (e.g. *alcoholic beverages*), as evidenced by a bias to respond "yes" with increasing phrase frequency. This suggests that phrase frequency *is* represented in long-term memory, either as a single coherent representation or as individual episodes.

Recognition memory data provide a perspective on how speakers of a language map between linguistic material and a context. A canonical view of recognition is that, at test, speakers are given the linguistic content, the test items, and have to retrieve the experimental context in which they were experienced in order to endorse the items as old (Reder et al., 2000). The demands of a recognition task are therefore more comprehension-like than production-like. The other major memory task, recall, works in the opposite way. An act of recall starts with an existing temporal, discourse, or situational context representation ("recall all of the words *on the list you just saw*") and maps to the linguistic material that was experienced in this context (Howard & Kahana, 2002; Criss, Aue, & Smith, 2011). Recall is an explicit language generation task. In this respect, the demands of recall are more akin to production than comprehension. The current studies therefore examine phrase frequency effects in recall, rather than recognition, to provide a different perspective on the question of the source of such effects and what they tell us about phrasal representation. The next section discusses findings on the production of adjective-noun phrases.

Many studies have found that language production is easier when participants have to produce frequent words and phrases. Word and phrase frequency effects are apparent in a number of production measures including faster onset times (Janssen & Barber, 2012) and shorter word durations in frequent phrases (Arnon & Cohen Priva, 2013; Bannard & Matthews, 2008). Janssen and Barber assessed whether phrase frequency as measured by hits on the Google search engine predicted how easily speakers provided modified noun phrase picture descriptions like *blue car* or *red house* and noun-noun pairs like *bus car* in Spanish as well as noun-adjective pairs in French. They measured speech onset latencies as a function of phrase frequency, the frequency of the first word, and the frequency of the second word in each pair. They found that phrase frequency, but neither of the individual word frequencies, predicted speech onset latencies. Generally, the higher the phrase frequency, the earlier speakers began talking. Because they found phrase frequency effects, Janssen and Barber argued that phrases are stored holistically and that these representations lack a relationship between the component words and the phrase.

The results of Janssen and Barber were surprising because a previous study by Alario, Costa, and Caramazza (2002) had identified separable contributions of adjective and noun frequency to speech onset latencies, where high frequency adjectives and nouns sped up noun phrase production. Janssen and Barber argued that the results of Alario et al. could have also been due to variations in phrase frequency confounded with word frequency, as high frequency phrases tend to be made up of high frequency words, which have well-known frequency effects. Even when controlling for word frequency, phrase frequency explained the ease of phrase production.

Additional evidence from child production data corroborates the hypothesis that the production system retrieves multiword units, perhaps in addition to individual words. Bannard and Matthews (2008) used a phrase imitation task in which children repeated phrases that an experimenter said to them. Children made fewer errors, and took less time to produce the overlapping words, when repeating more common phrases (e.g. "a drink of milk") than less common ones that shared the same first three words (e.g. "a drink of tea”). This suggests that long-term memory for multiword sequences has an effect on children’s language production.

Theories of language production have not had a great deal to say about the production of phrases, with the possible of exception of idiomatic phrases. The notion of a superlemma referred to earlier was developed by Sprenger et al. (2006) to allow for the model of Levelt, Roelofs, and Meyer (1999) to be able to produce idiomatic phrases. For non-idiomatic or compositional phrases, models have not assumed the existence of stored representations of multiword sequences (MacKay, 1982, is an exception in this respect). Because of the need for the production system to be able to assemble completely novel phrases (e.g. “an ugly beauty” cited by Chang, Dell, & Bock, 2006), models have emphasized that structural frames (e.g. adjective-noun) are retrieved, and then individual words UGLY and BEAUTY are retrieved and linked to slots in the frame (e.g. Chang et al, 2006; Dell, 1986; Dell, Oppenheim, & Kittredge, 2008; Garrett, 1975). Finding that production processes are sensitive to phrase frequency (e.g. Janssen & Barber, 2012; Bannard & Mathews. 2008) forces an amendment to these models.

To better understand phrase frequency effects, we consider the task of immediate free recall, which is an episodic memory task that engages the production system. We ask how phrase frequency supports retrieval for production. We will contrast phrase recall performance with recall of individual words. The first experiment (Experiment 1) explores the effects of word frequency on single-word (noun) recall, while Experiment 2 and Experiment 3 examine the influence of phrase frequency on recall of adjective-noun phrases.

Experiment 1

Frequency effects on free recall of nouns

The purpose of Experiment 1 is to examine whether a set of single words that show strong frequency effects in recognition in favor of the low frequency items (Balota, Burgess, Cortese, & Adams, 2002; Jacobs, et al., 2016) exhibit similar frequency effects in a free recall task. Some studies have found no effect of frequency on recall (Clark & Burchett, 1994; MacCleod & Kampe, 1996; Hulme et al., 2003), while others have found an advantage for high frequency words (Criss et al., 2011; Balota & Neely, 1980).

Methods

Materials

Study items were those used in Experiment 2 of Chapter 1. These items consisted of 88 nouns taken from Balota et al. (2002) that varied in frequency in the Google 1T n-gram corpus (Brantz & Franz, 2006), ranging from *parasol*, *sleuth*, and *crevice* on the low end to *car*, *book*, and *world* at the high end.

Participants

Thirty individuals from the University of Illinois paid subject pool received $8 for participating. All were native English speakers who acquired no language other than English before the age of 5.

Procedure

Participants carried out an immediate free recall test of four 22-word lists. Each was made aware prior to list study that immediate written recall would take place. Each participant saw a unique ordering of 88 nouns that were randomly assigned to four lists, with the additional constraint that each list contained 11 high and 11 low frequency words. Study order was randomized within the list. Every word was presented at the center of the computer screen for 1 second, followed by a 1 second inter-stimulus interval before the presentation of the next item.

After the end of the presentation of each list, the computer presented a prompt for participants to start recalling the words they studied on a piece of paper with 22 spaces for each list. The prompt said, "Please fill in as many of the words as you can remember in any order you would like. Please try to recall as many words as you can." After acknowledging the instructions, the screen displayed a countdown showing the remaining amount of time to recall that list (5 minutes per list was allotted). At the end the five-minute recall period, participants could initiate study of the next list when they wished to by pressing a key.

Results

Every word that participants wrote down was entered as a data point for analysis. If participants wrote down an item that had appeared on an earlier list, that item was considered an intrusion and excluded from analysis. If items were misspelled but sufficiently similar to be identified as another item on the list (e.g. "alter" for "altar" or "yach", "yaght" and "yatch" for "yacht"), that item was included. Items that were not on any list that participants studied were not considered in the analysis.

To analyze the effect of word frequency on word recall, we constructed a logit mixed model of whether each item that participants studied was recalled or not as a function of (log transformed) word frequency. Random effects of participant and item on the intercept and a random effect of participant on word frequency were included in the model.

Word frequency was not a significant predictor of the likelihood of the recall of a word. These results are summarized in Table 2.1 and plotted below in Figure 2.1.

	Estimate	SE	t
(Intercept)	-0.23	0.10	-2.19
(Log) phrase frequency	-0.05	0.05	-1.02

Note: Significance at |t| > 2.00

Table 2.1**:** Effect of word frequency on likelihood of noun recall.

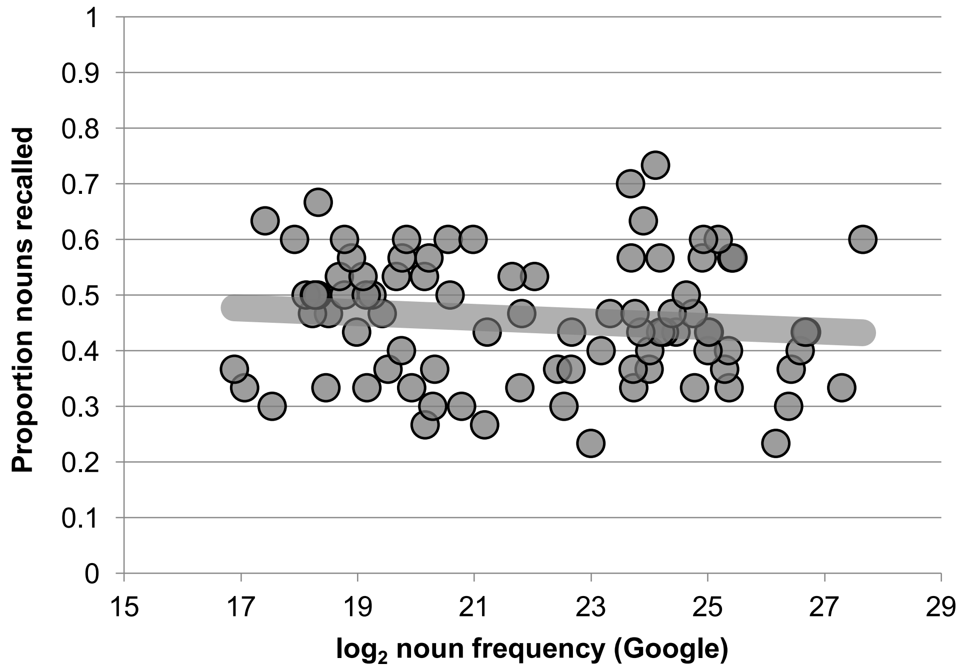

Figure 2.1: Effect of word frequency on free noun recall, Experiment 1. More common nouns like *tree* are recalled just as often as less common nouns like *wizard*.

Discussion

The results of this study replicate prior research findings of no high-frequency word advantage in the immediate recall of unrelated lists of nouns. Regardless of their individual frequencies, words are relatively equally well recalled across all frequency ranges. Apparently, the strength of associations from episodic context to items does not reflect the commonness of the words. By itself, this null result does not have strong implications about the nature of lexical storage and retrieval. As we will see, however, the findings of Experiment 1, known effects of word frequency on recognition memory, and the effects of phrase frequency on recall that we will report in Experiments 2 and 3 will provide useful constraints on a model of the representation and retrieval of words and phrases.

Interim Discussion

The results of Experiment 1 demonstrate that high frequency words are not necessarily better recalled than low frequency words. While this is in line with a number of previous studies that have not found an effect of word frequency on free recall, the pattern of results here differs from the expected pattern known to occur in less memory-focused language production tasks - when speakers are asked to name pictures, they are faster and more fluent in using high-frequency words (Jescheniak & Levelt, 1994; Dell, 1990; Kittredge et al., 2008).

The lack of a frequency effect on single-word recall can be explained if we consider the nature of recall. Recall can be conceptualized as a two-step process. First, recalling a word may involve mapping from the person's representation of the list of items they studied (hereafter known as the *episodic context*) to the word's semantic and/or syntactic representation, which is more formally known as the lemma. Second, once this representation is retrieved, the speaker must use the spoken or written production system to output the word.

The research mentioned above, taken together, specifically shows that it is the process of converting the lemma into speech or writing that is strongly sensitive to word frequency, rather than the retrieval of the lemma itself. The output of this conversion is ultimately phonological in nature, because speakers produce a sequence of sounds, letters, or characters. In unimpaired speakers this sensitivity is largely revealed in response time, rather than accuracy. Even though the word "wizard" is not particularly common relative to a word like "tree", when one has retrieved the lemma WIZARD, typical speakers accurately produce the word. By contrast, differences in production accuracy due to frequency only typically emerge in impaired populations (e.g. Kittredge et al., 2008), with one exception. Difficulty in phonological form retrieval during production is largely restricted to extremely low frequency words such "hemoglobin" or "ambergris." Difficulties retrieving these words often manifest as tip of the tongue states, where the sounds corresponding to the word cannot be retrieved from what the speaker means to say (e.g. Brown & McNeill, 1966; Rubin, 1975; Harley & Brown, 1998).

Given these considerations, word frequency should not impact the production component of a typical untimed free recall task in which the words are known to the participant. Thus, if Experiment 1 had shown a substantial word frequency effect in free recall, it would have demonstrated frequency sensitivity in the link from episodic context to lexicon. Given that we and many others found no such word frequency effect on free recall, we tentatively conclude that word frequency is not a powerful influence on the episodic retrieval of a word.

Should phrase frequency then also not matter in free recall? One potential mechanism underlying the episodic retrieval of phrases is that phrase production benefits from pattern completion, otherwise known as *redintegration*. During redintegration, long-term memory associations between components of a to-be recalled item help to fill in the gaps in memory when not all components are initially retrieved (Schweikert, 1993; Horowitz & Manelis, 1972). Phrase recall importantly differs from word recall in that phrases, unlike most words, are systematically composed of meaningful components (i.e. words). To the extent that free recall is driven by the retrieval of meaning (Hill, Jones, & Todd, 2012), one would expect systematic incomplete or partial recall in which some words are correct but not others. In such a case redintegration would mean that recall of some words of a phrase may help a speaker retrieve the other words. This process of redintegration may be sensitive to phrase frequency.

To see how phrase frequency might matter in recall, let us be more specific about redintegration in the recall of an adjective-noun phrase such as "alcoholic beverages." Assume that when the retrieval process starts, there is some probability p that at least one of the words is retrieved. Then, as recall continues, the process may succeed in full recall with probability q. Alternatively retrieval may fail, leading to the correct recall of only one word from the studied phrase. These possibilities are illustrated in Figure 2.2. One can use this simple model to derive expectations about the role of phrase frequency in recall. Would high phrase-frequency aid initial recall, parameter p of the process? If we assume that initial recall is driven largely by the strength of the episodic associations from the list context to the language system and that these associations are not sensitive to frequency, as we claimed for single-word recall, then we do not expect a consistent effect of frequency on this parameter. Because phrases, however, are systematically composed of meaningful parts, retrieval from long-term memory representations may take advantage of connections between these components via a *redintegration* process. If so, we would expect more common phrases to be associated with *complete* recall, that is, to have a larger value of q (complete recall given some recall). We will postpone a consideration of specific mechanisms for such a process until we gather new data.

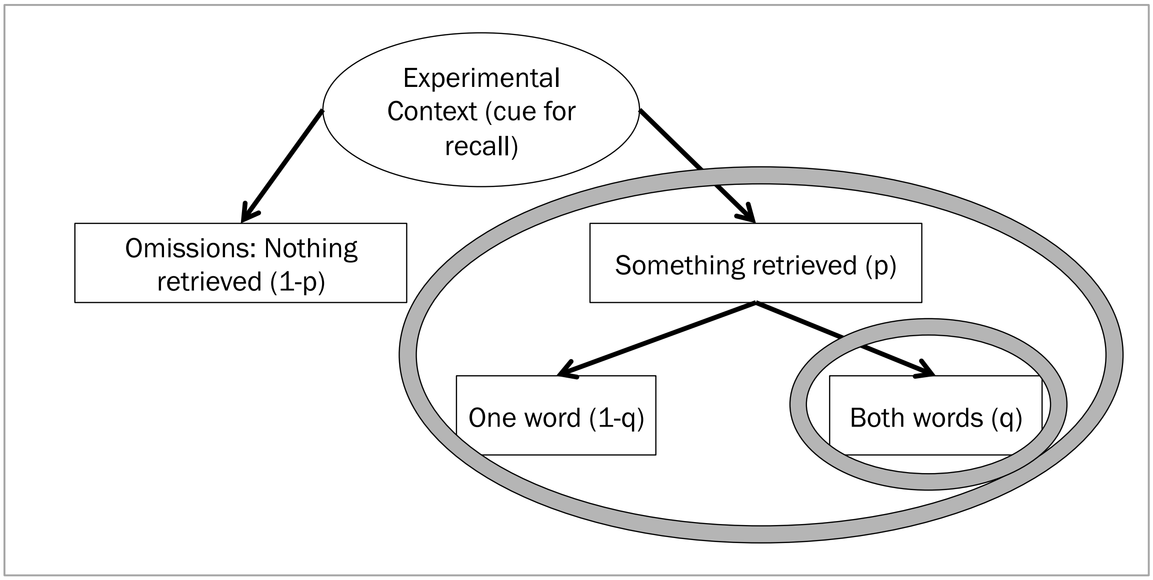

Figure 2.2: In order to identify whether phrase frequency has an effect on the recall of phrases, two parameters can be estimated. First is the *p* parameter, which measures the likelihood of recalling at least one part or the whole phrase (that is, either "alcoholic", or "beverages", or "alcoholic beverages") about a phrase versus recalling nothing about a phrase. Second is the *q* parameter, which measures like likelihood of recalling the entire phrase (i.e. "alcoholic beverages") given that a phrase has been recalled.

To test these proposals, Experiments 2 and 3 use recall tasks structured similarly to Experiment 1, except that the stimuli are meaningful adjective-noun phrases. In Experiment 2a, participants are presented with adjective-noun phrases designed to vary only in phrase frequency. After receiving the last phrase in a study session, participants must recall the phrases by writing them down. Experiment 2b is a replication, except that participants are told to recall the individual words. This change was implemented in order to see whether the phrasal organization at output influences any phrase frequency effects. In Experiment 3, we sought to see whether the results of Experiment 2 generalized to another set of phrases and a different procedure in which study time per phrase was determined by the participants rather than being experimenter controlled. The results of these experiments allow us to test whether phrase frequency effects arise at initial recall of a phrase (complete or incomplete), or only after a participant has already recalled one of the words of a phrase.

Experiment 2a

Frequency effects on immediate free recall of adjective-noun phrases

If phrases are processed and remembered just as big words, then we expect phrase recall to be unaffected by phrase frequency, as seen in Experiment 1 with individual words. Hence, Experiment 2 looks at the effect of phrase frequency on free recall of phrases. Critically, phrase recall is prone to errors that single words cannot generate: parts of phrases can be recalled. We can capture this by estimating the two parameters that we outlined earlier, probability of some recall (p) and probability of complete recall given some recall (q).

Methods

Materials

Phrases from this experiment were a subset of the 112 phrases used in Experiment 3 of Chapter 1. These phrases were taken from the Corpus of Contemporary American English (COCA; Davies, 2008) and included items such as "critical condition", "horrible mistake", and "impossible dream." To ensure that our assessment of the influence of phrase frequency on recall was not the result of any confounding between frequency and compositionality or concreteness, we conducted a norming study on Qualtrics in which

University of Illinois undergraduates rated the items along several dimensions and completed the questionnaire at their own pace.

In this norming study, 30 participants provided responses to a number of questions on a five-point Likert scale from "Strongly Disagree" to "Strongly Agree". First, familiarity with the component words of each phrase and the phrase itself was assessed; participants answered whether they knew the meanings of, for example, the word "impossible", "dream", and the phrase "impossible dream." Then, to rate the imageability of the phrase, participants rated whether they could easily picture what this phrase describes. Finally, as a measure of compositionality, participants rated whether "impossible dream" had the same meaning as a dream that is impossible. Ratings were averaged across all participants and then centered and scaled with respect to all items for inclusion in the analyses. In the final stimulus set, phrases were restricted to just those where the average imageability and compositionality scores fell within a narrow range in order to decorrelate imageability and compositionality from phrase frequency ($r = .11$, $t(70) = 0.89$, $p =$ n.s. for imageability; $r = -.14$, $t(70) = -1.19$, $p =$ n.s., for compositionality). After these requirements were met, 72 phrases remained. These stimuli are available in Appendix D.

Participants

In the norming study described above, 30 undergraduate students from the University of Illinois were recruited from the course credit subject pool. All participants were native speakers of English who acquired no language before the age of 5. Each participant received one hour of credit for participation.

For the memory component of this study, 40 undergraduate students recruited from the University of Illinois course credit subject pool participated in this experiment with the same qualifications as the norming study. Each person received one hour of credit for participation in the experiment.

Procedure

The 72 items were broken into 4 lists and were randomly populated in the same way as in Experiment 1. Each list contained an equal number of high and low frequency phrases. For each list, participants studied 18 phrases for 1.5 seconds each followed by a 1 second inter-trial interval. After studying the 18th phrase, participants were told, "Try to write down as many of the phrases as you can remember. If you cannot remember both of the words from a phrase, but just one of the words, then write that down instead." Participants were given 5 minutes to complete recall of each list, again with a countdown informing them about how much time was left. If participants finished ahead of time, they waited until the timer finished before beginning study of the next list.

Results

Scoring

Each recalled item was categorized for whether the adjective was correctly recalled, whether the noun was correctly recalled, or both, as well as in what position in the recall list participants recalled the whole phrase or only part of the phrase. As before, items that could be identified as the target based on misspelling were included as correctly recalled in the analysis.

Phrase recall can be conceptualized as a two-stage process (e.g. Schweickert, 1993), which is summarized graphically below in terms of the parameters p and q in

Figure 2. Participant responses on each individual trial were coded in terms of these whether participants had recalled at least one word ($p = 0$ or 1), and if they had recalled at least one word, whether they had recalled just one or both ($q = 0$ or 1). We then performed a sequential logistic regression analysis (e.g. Fox, 1997), fitting independent binary logistic models to each of the two stages. This tells us about the effect of phrase frequency on the likelihood of recalling anything from a phrase (first analysis) and the likelihood of partial versus complete recall (second analysis) respectively (that is, the p and q parameters).

Mixed effects logistic regression models were built to test for the effect of phrase frequency, (quadratic) study order (typical for allowing primacy and recency effects, e.g. Freebody & Anderson, 1986; Anderson & Bower, 1972; May & Sande, 1982), and concreteness on the likelihood of first some (p) and then complete-given-some (q) recall. Random effects were the participant-level random intercepts and random slopes of phrase frequency, with random intercepts by item.

Similar to the pattern of results in Experiment 1, where word frequency did not influence single word recall for nouns, we found that phrase frequency did not influence the likelihood at least one word of a phrase being recalled (the p parameter). Concrete phrases like "private plane", however, were more likely to be recalled at least in part than more abstract phrases like "critical condition." This occurred in spite of the relatively narrow range of concreteness values. Additionally, there was the expected effect of study order, with expected serial position effects as seen in the significant quadratic study order term. These results are summarized below in Table 2.2.

	Estimate	SE	t	
(Intercept)	-0.41	0.11	-3.55	***
Study order (quadratic)	2.22	0.19	11.59	***
(Log) phrase frequency	0.01	0.10	0.12	
Phrase concreteness	0.32	0.10	3.17	**

Note: Significance at |t| > 2.00

Table 2.2: Effect of phrase frequency on parameter p, the recall of adjective-noun phrases, COCA stimuli, Experiment 2a. More common phrases are as likely to be recalled at least in part as less common phrases, but concrete phrases are more likely to be recalled.

The results with parameter q were different. There was a significant positive relationship between phrase frequency and the likelihood of the phrase being recalled in its entirety (given recall of at least one word) - high phrase frequency helped participants produce both words from studied phrases. More concrete phrases were also more likely to be recalled in their entirety. These results are summarized below in Table 2.3. Both analyses are plotted in Figure 3 below.

	Estimate	SE	t	
(Intercept)	1.92	0.21	9.28	***
Study order (quadratic)	1.21	0.39	3.10	**
(Log) phrase frequency	0.34	0.16	2.03	*
Phrase concreteness	0.50	0.15	3.38	***

Note: Significance at |t| > 2.00

Table 2.3: Effect of phrase frequency on parameter q, the complete versus incomplete recall of adjective-noun phrases, COCA stimuli, Experiment 2a. More common phrases are more likely to be recalled in their entirety than less common phrases.

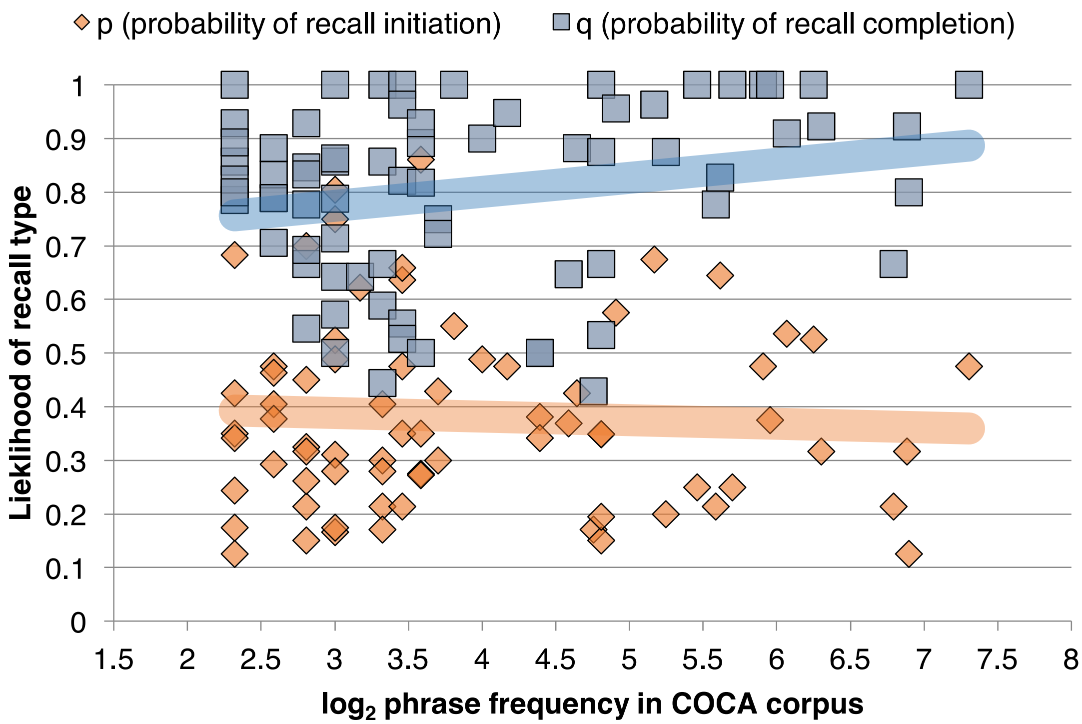

Figure 2.3: Effect of phrase frequency on the recall of adjective-noun phrases from COCA, Experiment 2a. More common phrases are more likely to be recalled in their entirety (blue squares) than less common phrases, but all are equally likely to be recalled at least in part (orange diamonds).

Experiment 2b

Experiment 2b was a replication of 2a with a change to recall instructions, emphasizing recall of words, rather than recall of phrases. Participants studied the same phrases as in Experiment 2a, but were told to write down as many of the individual words as they could remember.

In addition to providing information about how easily the *words* within phrases are encoded and retrieved, this experiment can show how much phrases can *incidentally* be reconstructed from long-term memory even when the task is to recall words, and not pairs of words. If people write down both words from a phrase more often as a function of studied phrase frequency, then this is evidence that phrase structure is an important incidental organizing feature. If it is, one can then ask what the role of phrase frequency is in the same manner that we did for Experiment 2a.

Methods

Materials

Materials were the same as those from Experiment 2a.

Participants

40 participants from the University of Illinois course credit or paid subject pool took part in this experiment. All participants were native English speakers who acquired no other language before the age of 5.

Procedure

The study procedure of this experiment was identical to that of Experiment 2a. The recall phase differed in the instructions given to the participants about the nature of their responses after study. Participants were told, "You are going to see a series of two-word phrases presented on the screen. While they are two words presented together, we want you to remember each of the individual words separately because you will be asked to write down the individual words on separate lines from memory. If you remember both words from a phrase, write each word on a separate line." After participants began the test phase, they again had 5 minutes to recall as many of the words as possible by writing their answers on sheets of paper with provided spaces. At the end of the five-minute recall period, participants pressed a key to begin the next study-test phase.

Results

Experiment 2b replicated the effects of Experiment 2a. Participants wrote down at least one word from a phrase as often across all frequency ranges (results for parameter p in

Figure 2.4 and Table 2.4), but were significantly more likely to recall both words from high frequency phrases given recall of at least one word (results for q in Figure 2.4 and Table 2.5). Words from more concrete phrases were more likely to be recalled (p) and were more likely to be recalled if their phrasal mate had been recalled (q). Finally, as before words that had occurred in phrases early or late in the list were better recalled than words from phrases in the middle of the list.

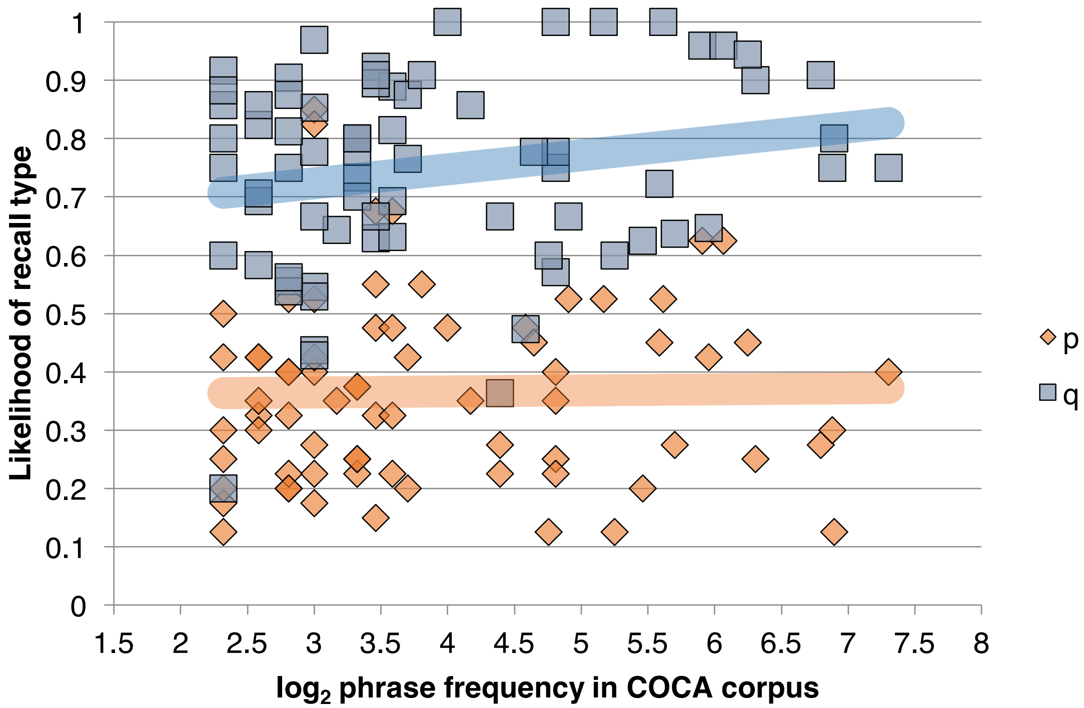

Figure 2.4: Effect of phrase frequency on the recall of individual words from adjective-noun phrases from COCA, Experiment 2b. More common phrases are more likely to lead to both words being recalled (blue squares) than less common phrases, but all are equally likely to be recalled to some extent (orange diamonds).

	Estimate	SE	t	
(Intercept)	-3.86	0.62	-6.14	***
Study order (quadratic)	0.01	0.001	8.60	***
(Log) phrase frequency	0.03	0.03	1.04	
Phrase concreteness	0.31	0.07	4.78	***

Note: Significance at |t| > 2.00

Table 2.4: Effect of phrase frequency on parameter p, the recall of any versus all words from adjective-noun phrases, COCA stimuli, Experiment 2b.

	Estimate	SE	t	
(Intercept)	-0.99	0.10	-9.86	***
Study order (quadratic)	0.40	0.04	9.10	***
(Log) phrase frequency	0.13	0.04	2.76	**
Phrase concreteness	0.23	0.05	5.13	***

Note: Significance at |t| > 2.00

Table 2.5: Effect of phrase frequency on parameter q, the complete versus incomplete recall of adjective-noun phrases, COCA stimuli, Experiment 2b. More common phrases are more likely to be recalled in their entirety than less common phrases.

Discussion

Experiment 2b combined features of Experiment 1 and Experiment 2a by assessing whether the recall of individual words from concrete, compositional phrases was affected by phrase frequency. Like words, compositional phrases were recalled at least in part (parameter p) equally well at all levels of the frequency range, which is similar to the effect found in Experiment 1 for individual words. Experiment 2a found that once something had been retrieved from a phrase, though, the phrase was more likely to be completed if it was a high frequency phrase than if it was a low frequency phrase (parameter q). Experiment 2b replicated these results, demonstrating that long-term memory representations of high frequency phrases are useful for pattern completion, in that the retrieval of one word in a phrase facilitates the retrieval of the other word in a phrase.

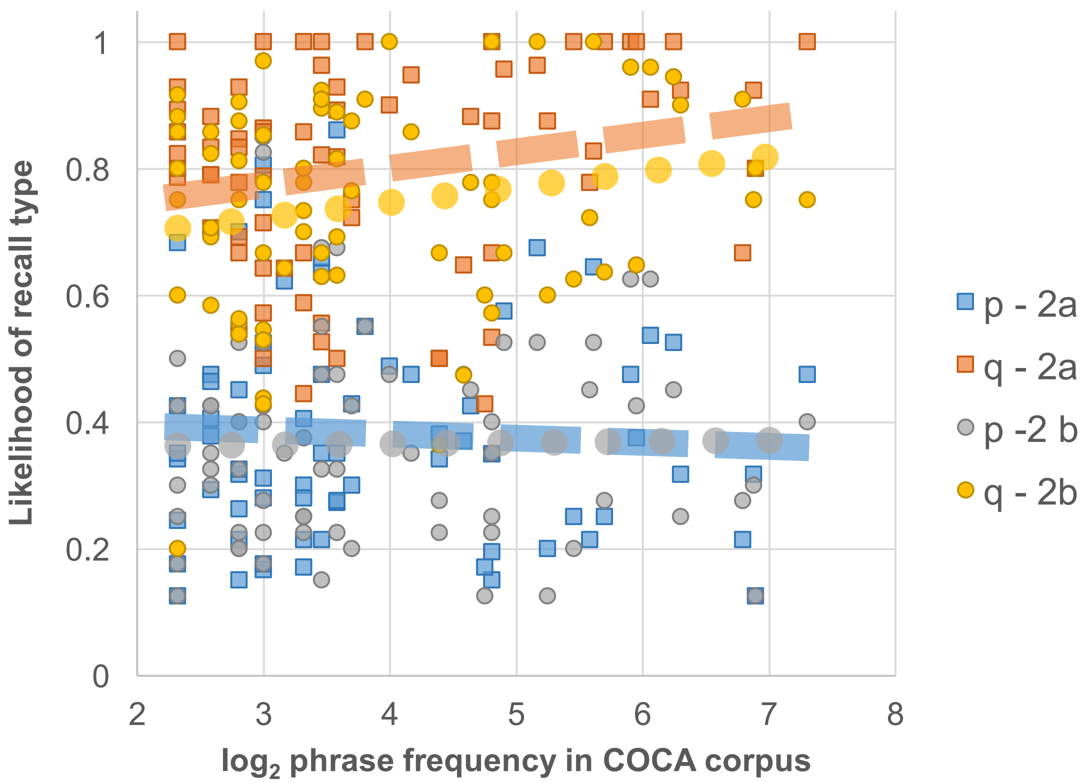

Figure 2.5: Likelihood of complete recall of phrases given recall of at least one of the words as a function of phrase frequency. The effect of phrase frequency is similar in both Experiment 2a and 2b for both the likelihood of remembering either one or two words (p) and for the likelihood of remembering two words when at least one word was recalled (q). Participants are less likely to recall phrases completely in Experiment 2b, when they were prompted to only write down words.

There was one difference in the results of Experiments 2a and 2b (see Figure 2.5). When participants recalled an item, they recalled both words of the phrase as opposed to just one word in Experiment 2a on average 79.8% of the time, while in Experiment 2b this value was only 74.4%. In a paired t-test comparing p and q parameters of the two experiments, although the effect of frequency on q was similar in the experiments, the q values themselves are significantly lower in Experiment 2b ($t(71) = -2.27$, $p = .013$). At the same time, participants were just as likely to recall an item in whole or in part (parameter p) in Experiment 2b as 2a ($t(71) = -0.92$, $p = 0.185$).

It is striking that even when the task is not to recall phrases, but instead individual words, the influence of phrase frequency on word recall is similar to its effect in phrase recall. This suggests that phrasal organization in long-term memory is the driving force behind phrase frequency effects in free recall. Furthermore, there is a dissociation between p and q in how influential the instructions are. Telling participants to write down single words as opposed to phrases affects the likelihood of participants writing down both words of a phrase when they recall an item (q), but does not influence the likelihood of them recalling at least one word from that item (p).

In summary, even though participants were asked to recall individual words, the task demands did not prevent them from recalling both words from a phrase. This is consistent with the finding that in single word recall, participants often attempt to recall temporally contiguous or semantically related words at the same time (Unsworth, Brewer, & Spillers, 2014; Gruenewald & Lockhead, 1980; Wixted & Rohrer, 1994; Sederberg, Howard, & Kahana, 2008; Lohnas & Kahana, 2014). The results of Experiment 2 are consistent with the idea that the initial recall of a word or phrase is insensitive to phrase frequency, but that once a part of the phrase has been recalled, phrase frequency becomes an important catalyst in recalling an entire phrase.

Experiment 3

Frequency effects on self-paced study and free recall of adjective-noun phrases

Experiment 2 demonstrated that phrase frequency can affect aspects of phrase recall, especially during the process of completing recall of an entire phrase. Experiment 3 aimed to replicate and extend the phrase frequency effects of Experiment 2 in a recall paradigm where participants can pace their own study and where the materials differ from prior materials by having a wider range of concreteness scores. While it was less clear what would happen in more natural materials with the likelihood of the initial recall of any given phrase (the p parameter), the analysis of the q parameter representing the likelihood of redintegration remains the critical analysis. If phrase frequency influences the likelihood of the complete recall of a phrase, then Experiment 3 should replicate the effects of Experiment 2 on the q parameter, with high frequency phrases being more likely to be recalled in their entirety than low frequency phrases.

Methods

Materials

Experiment 3 used the 52 phrases from Experiment 1 of Chapter 1 as stimuli such as “alcoholic beverages” and “psychic nephew”. These items varied in their phrase frequency, which was decorrelated by design from adjective frequency, noun frequency, and both word lengths, but which somewhat confounded concreteness with phrase frequency (see Chapter 1 for details). Phrase frequency and concreteness were correlated ($\rho = .49$), which we account for in later analyses by performing likelihood ratio tests.

Participants

Seventy-nine undergraduate participants were recruited from the University of Illinois course credit subject pool. All participants were native speakers of English who acquired no language other than English before the age of 5. Each person received one hour of credit for participation in the experiment.

Procedure

Each participant did two study-test blocks of 26 phrases each that were randomly populated in the same way as in Experiments 1 and 2. Phrases were presented at the center of the screen until participants pressed the space bar to continue on to the next phrase followed by a one second inter-item interval. After pressing a key to complete study on the 26th item, the test phase began. Participants were told, "Try to write down as many of the phrases as you can remember. If you cannot remember both of the words from a phrase, you can write down just one of the words." Participants were given 10 minutes per list to recall as many of the items as they could remember by writing their answers on sheets of paper with provided spaces. Participants waited the entire interval before beginning the second study-test phase.

Results

Random effects were structured in the same way as Experiment 2. Fixed effects of interest included how long a participant studied each item in log seconds, quadratic study order, the concreteness of each phrase taken from the norms of Chapter 1, and the log frequency of the phrase. Because concreteness and study time were somewhat confounded with the variable of interest (phrase frequency), we performed likelihood ratio tests for whether including phrase frequency in the model explained variance over above that explained by a model containing only study time, concreteness, and study order. When the likelihood ratio test revealed that adding frequency gave a significant improvement in fit, we included phrase frequency in the final model.

Due to the self-paced nature of the task, we were interested in whether participants studied phrases more when they were infrequent, which could weaken or eliminate any phrase frequency effects on memory (though see the *laboring in vain effect*; Nelson & Leonesio, 1988). Participants indeed labored longer on less common phrases (B = -0.013, t = -4.57, $p < .001$), in line with similar frequency-related processing fluency gains in studies of language comprehension (Smith & Levy, 2013; Arnon & Snider, 2010; Siyanova-Chanturia et al., 2011). As we show below in Figure 2.6, though, they labored in vain, as more common phrases were still better recalled.

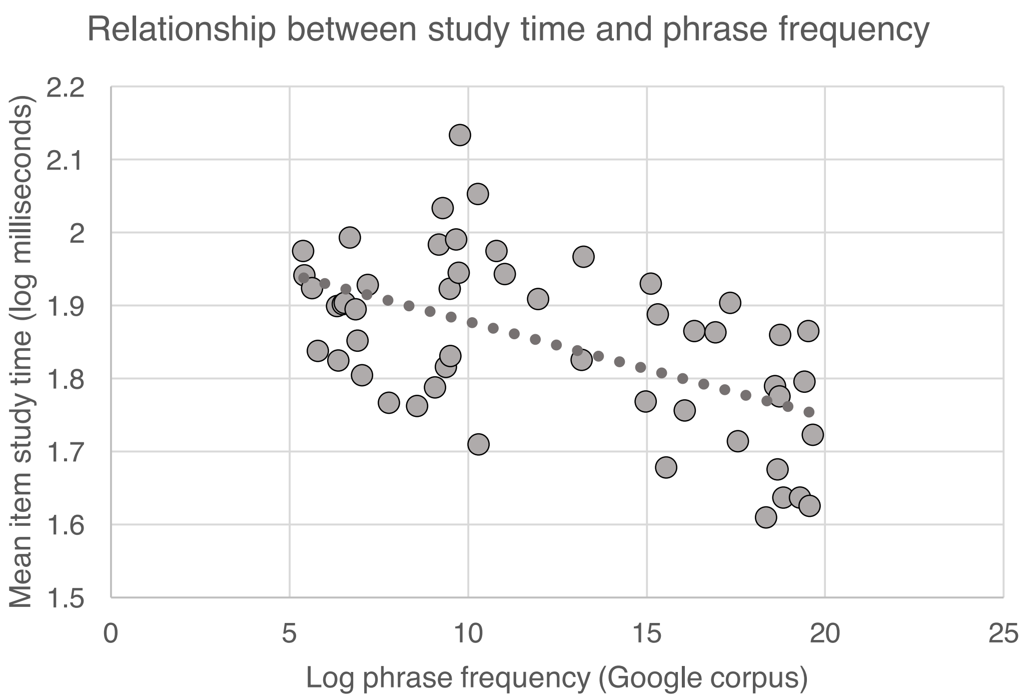

Figure 2.6: Effect of phrase frequency on study time. More common phrases are studied for less time.

The first analysis focuses on the likelihood of recalling at least one word from a phrase (parameter p). The analysis showed that length of time the participants studied an item, the order of the item in a list, and its concreteness all influenced the likelihood of it being recalled, either in whole or in part. The inclusion of phrase frequency improved model fit beyond these control variables ($\chi^2(1) = 8.46$, $p < .01$). The model assessing the effect of phrase frequency on the p parameter is summarized below in Table 2.6.

	Estimate	SE	t	
(Intercept)	-0.38	0.13	-2.82	**
(Log) phrase frequency	0.35	0.12	3.03	**
(Log) study time	0.51	0.07	7.67	***
Study order (quadratic)	0.40	0.05	8.51	***
Phrase concreteness	0.30	0.12	2.48	*

Note: Significance at $|t| > 2.00$

Table 2.6: Effect of variable on the p parameter of the recall of adjective-noun phrases, Google stimuli. Concrete phrases, those that are studied for longer, and higher frequency phrases are more likely to be recalled.

Focusing on the q parameter, phrase frequency importantly continued to have an effect on recall performance. The model containing phrase frequency, concreteness, study time, and study order was a better fit than a model containing only those other factors in a likelihood ratio test, so we report the larger model for the likelihood that a participant will recall a full phrase, rather than a partial phrase ($\chi^2(1) = 6.99$, $p < .01$). Even when controlling for concreteness and study time, higher frequency phrases like "alcoholic beverages" were more likely than lower frequency phrases like "psychic nephew" to be recalled as wholes. Phrases studied for longer as well as those with higher concreteness ratings were more also associated with higher values of q. These results are summarized below in Table 2.7.

	Estimate	SE	t	
(Intercept)	2.14	0.22	9.70	***
(Log) phrase frequency	0.57	0.21	2.66	**
(Log) study time	0.29	0.12	2.37	*
Study order (quadratic)	0.07	0.10	0.70	
Phrase concreteness	0.44	0.21	2.04	*

Table 2.7: Effect of phrase frequency on the complete versus incomplete recall of adjective-noun phrases, Google stimuli. More common phrases are more likely to be recalled in their entirety than less common phrases.

Below are plotted the likelihood of complete recall given any recall as a function of phrase frequency in Figure 2.7.

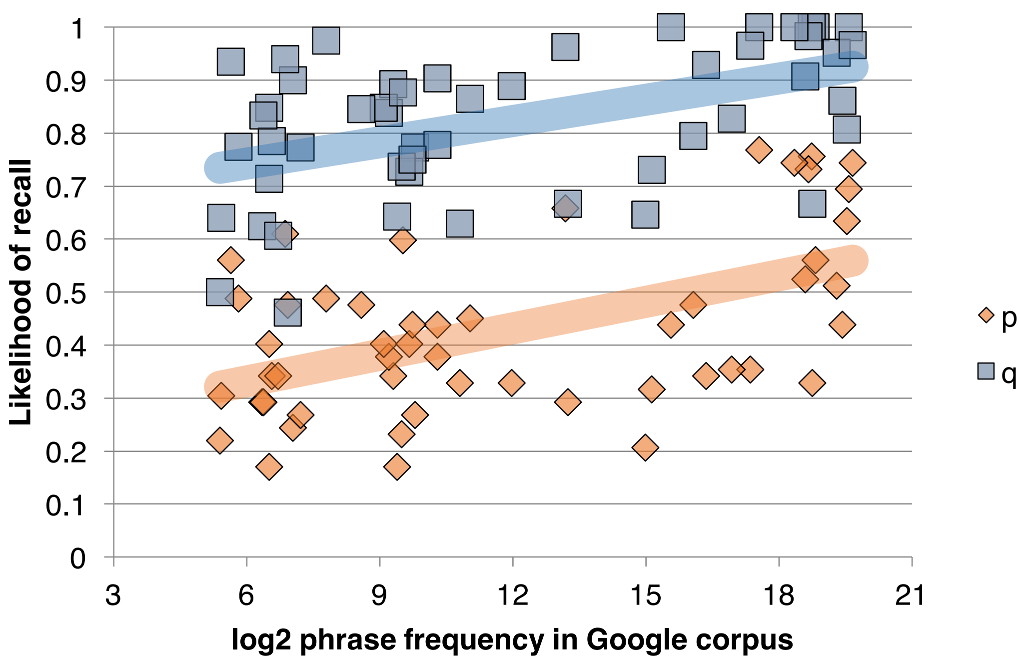

Figure 2.7: Effect of phrase frequency on the likelihood of complete (as opposed to incomplete) recall of adjective-noun phrases, Google stimuli, Experiment 3. More common phrases are more likely to be recalled in whole or in part than uncommon phrases, and are more likely to be recalled in their entirety given that at least one of the words was recalled (blue squares) than less common phrases.

Discussion

Experiment 3 demonstrated that phrase frequency has a strong influence on the likelihood of a phrase being recalled in its entirety, given some recall, replicating the findings of Experiment 2. Experiment 3's replication of Experiments 2a and 2b's positive phrase frequency effect on q solidifies a conclusion that redintegrative processes drive the reproduction of phrases from memory. A possible explanation of this effect is that phrasal representations consist of their constituent words, with some kind of link, such as a direct association or a chunk node joining them. In any event, the phrases are not atomic. In the general discussion we consider these results in concert with other findings regarding phrase and word frequency effects in recall and recognition. We will also consider the presence of an effect of phrase frequency on parameter *p* in this experiment, unlike in Experiments 2a and 2b.

General Discussion

Frequent linguistic units facilitate fluent language production. High frequency words are produced more quickly (Ellis, 2002; Gahl, 2008; Forster & Chambers, 1973) and are less prone to errors (Dell, 1990; Nozari et al., 2010). Production is a component of verbal free recall, so we can ask whether common linguistic units benefit in recall as well. Although the present study did not consistently find that phrase frequency contributed to the probability that at least one word of a studied adjective-noun phrase is recalled, the facilitative effect of frequency did show up as a greater likelihood of complete phrase recall (as opposed to partial recall). We characterized this finding as phrase frequency consistently affecting one parameter (q), but not the other (p), of a two-stage description of phrase recall.

The results of our experiments fit nicely with other investigations of the recall of of adjective-noun phrases (e.g. Horowitz & Manelis, 1972; Bower, 1969; Paivio, Khan, & Begg, 2000). In a seminal study, Horowitz and Manelis (1972) tested for the influence of idiomaticity on the free recall of these phrases. Phrases were either idiomatic expressions like *sour grapes*, meaningful (compositional) adjective-noun phrases like *green grapes*, or anomalous like *deep grapes*. Participants were told to write down as many phrases from memory as possible as part of a free recall task. As in the present study, Horowitz and Manelis were interested in whether the different kinds of phrases were more likely to be recalled as wholes, as opposed to partially. The tendency for complete recall was particularly strong for idioms, but it was nonetheless strongly present for all phrases, even novel phrases that were effectively meaningless (e.g. *deep grapes*). This effect demonstrates the influence of redintegrative processes during phrase retrieval.

Redintegration refers to a process of pattern completion using information from long-term memory (Horowitz & Manelis, 1972; Thorn, Gathercole, & Frankish, 2005; Schweickert, 1993; Hulme et al., 1997). We propose that specifically in phrase recall, the lexical/semantic representations of words that are retrieved during language production (e.g. lemmas and/or lexical concepts, Levelt et al., 1999) cue one another to the extent that they have often co-occurred.

Conclusion

We have examined word and phrase frequency effects in free recall. As is also true for such effects in recognition, the results are not straightforward. Words and phrases are not necessarily better recalled when they are more frequent. But in the case of phrases, there is a clear benefit for high frequency phrases for complete, as opposed to partial recall. We presented an informal model of these data and corresponding data in word and phrase recognition that put effects of word and phrase frequency in two locations in the cognitive system – within the lexical-semantic system that is responsible for language production and comprehension, and in the system that creates episodic memories based on the features that the lexical semantic system generates.

CHAPTER 3: MODEL

The following proposal outlines a model of our findings concerning the role of frequency in word and phrase recall. This model also explains the effects of word and phrase frequency in recognition memory (Chapter 1), who investigated phrase and word frequency effects using similar materials to those employed here.

The main challenge for a model of memory for linguistic material such as words and phrases is the fact that frequency effects appear to behave quite differently in recall and recognition. In particular, such a model must first be able to explain the well-known finding that more common words have considerably worse discriminability in recognition (Glanzer & Adams, 1985; Chapter 1, Experiment 3a), but, in single-word free recall, word frequency often has little impact on performance (our Experiment 1; Dunlap & Dunlap, 1979; Ozubko & Joordens, 2007). The results for phrases are even more complex, with frequency mattering for some aspects of each memory task, but not for other aspects. High frequency phrases are more likely to be recalled in their entirety once recall of a single word has been initiated (the consistent effects of phrase frequency on the q parameter), but there is a relative lack of phrase frequency effects on the p parameter, (Experiments 2a, 2b). In recognition, high-frequency phrases garner more "yes" responses during recognition tasks (Chapter 1, Experiments 1 and 2), but phrase-frequency does not impact actual discriminability. Instead, the frequency that impacts phrase discriminability in recognition is *word* frequency, specifically the frequency of the noun in adjective noun phrases (Chapter 1).

Finally, it is worth noting a property of phrase memory that appears to work similarly in recall and recognition: Concrete phrases are better remembered (Experiment 2a, 2b, and 3 for recall and Experiment 1 in Chapter 1; Kusyszyn & Paivio, 1966; Paivio et al., 2000). In Table

3.1, we summarize the pattern of results from the present word and phrase recall studies as well as the word and phrase recognition studies of Chapter 1, and provide a brief characterization of how each effect is explained in the model that we detail below in Figure 3.1.

Recall (Chapter 2)

Experimental result	Mechanism
Low frequency words and high frequency words are equally likely to be recalled.	Links from episodic context to lexical or semantic representations of words are independent of frequency
Low frequency phrases and high frequency phrases are under some conditions, equally likely to be recalled at least in part	Links from episodic context to lexical or semantic representations of words (and therefore phrases) are independent of frequency
High frequency phrases are more likely than low frequency phrases to be completed once one word has been recalled	Associations between the words within the lexical-semantic system are stronger in high frequency phrases
Concrete phrases are easier to recall than abstract phrases	Concrete phrases have more active features, so the associations between a new episode and a concrete phrase is stronger

Recognition (Chapter 1)

Experimental result	Mechanism
Low frequency words are better discriminated than high frequency words	Studied high frequency words suffer from more interference from prior episodes
High frequency phrases get more "yes" responses regardless of whether they were studied or not (a bias)	Associations between the words within the lexical-semantic system are stronger in high frequency phrases, contributing to greater familiarity
High and low frequency phrases are equally well discriminated	There are many more episodes sharing a word in a phrase than the whole phrase. Thus, interference from other phrase episodes is minimal.
Low frequency words facilitate phrase discrimination	Compositional phrases access word episodes, so high frequency words within phrases generate more interference just as they do in recognition for single words
Concrete phrases are better discriminated than abstract phrases	Concrete phrases have more active features, so the associations between a new episode and a concrete phrase are stronger

Table 3.1: Pattern of results that the model must be able to account for and proposed mechanisms

The model we propose combines features of language production models with prominent models of episodic memory (e.g. Reder et al., 2000; Howard & Kahana, 2002). An episodic memory is a link between features of the context and features of an item. The context represents the participant's surroundings, her internal state, and her conception of the task. During the study of a list, the set of context features will gradually change, but we assume that a great many will remain constant and thus represent the "list". The study item has features that represent properties of particular studied words and phrases. These features arise from processing the linguistic material using the lexical-semantic system that is used for language production and comprehension. Item features would include semantic, syntactic, and lexical properties of the item, as well as possible sensory-motor features that are called to mind by processing the material. In our model, we will represent the collection of features associated with a word, such as *cat*, by a single node. But this is a shorthand for the word's many properties. Furthermore, we make no claims about conditions that may favor more or less encoding of word-meaning features as opposed to word-form features, while recognizing that most studies of long-term memory like ours emphasize the encoding of meaning.

Recall and recognition are handled differently by the model but make use of the same architecture. The start point for recall is always the context, and the goal of recall is to use the context to retrieve linguistic material associated with it; that is, speakers are attempting to produce a word or a phrase. Recognition, instead, starts with the linguistic material as a cue. The recognition process succeeds (or generates a hit) when the linguistic input cues retrieval of the crucial experimental episode in which the material was studied. At the same time, the recognition process is influenced by the familiarity of the linguistic information, so unstudied material that is very familiar can trigger a “yes” response.

We assume that studied words and phrases are features of stored episodes. An episode is node connecting a representation of the episodic context and the lexical/ semantic representations of the linguistic material. The strength of the link between the context and the linguistic material is not assumed to reflect frequency of usage of the linguistic material. But material that is more concrete is assumed to contain more features and thus to have a potentially richer linkage.

More frequent words and phrases are assumed to be linked to more episodes. In addition, more common words have stronger connections to their phonological forms (e.g. Nozari et al., 2010; Kittredge et al., 2008; Jescheniak & Levelt, 1994; Dell, 1990). Phrases that have been heard or produced before include a link (or node, e.g. MacKay, 1982) connecting the lexical/semantic representations of their component words, with more common phrases having stronger connections.

These assumptions are illustrated in Figure 3.1. The pool of episodes contains numbered nodes that represent experiences. Each episode is therefore an instance, or exemplar, of a particular (potentially linguistic) category or combination of categories. In the figure, for example, Node 33 denotes a memory involving something big and something about cats, such as the phrase *big cat*. Likewise, Node 18 indicates an encounter involving something sad and a pug, potentially a *sad pug*. The lines linking lexical-semantic information to episodic events do not reflect frequency, but potentially concreteness and the activation or amount of attention devoted to the words.

The episodes are not all attached to exactly the same context features, since experiments unfold over time. A participant’s experience of the beginning of the experiment may be different from the end of it, for example. So, episodes should be able to be bound to different parts of a context. To illustrate this, the *big* and *cat* episode (Node 8104) connects to a different part of the context than the *sad* and *pug* episode (Node 8103), as different information may have been salient at time points 8103 and 8104.

Below we outline how these assumptions explain the word and phrase frequency effects in recall and recognition.

Recall

Low frequency words and high frequency words are equally likely to be recalled. This suggests that the long-term memory encoding process, that is, the linkage between each episode and the words that participated in it, is largely independent of any frequency information that is stored with the representations of words in the lexical/semantic network. Lexical frequency is presented in the lexical semantic system, but it is most strongly felt in the mapping from semantic/syntactic representations to phonological forms. During a recall test of familiar words by unimpaired speakers, the sensitive component of the mapping does not generate any appreciable error.

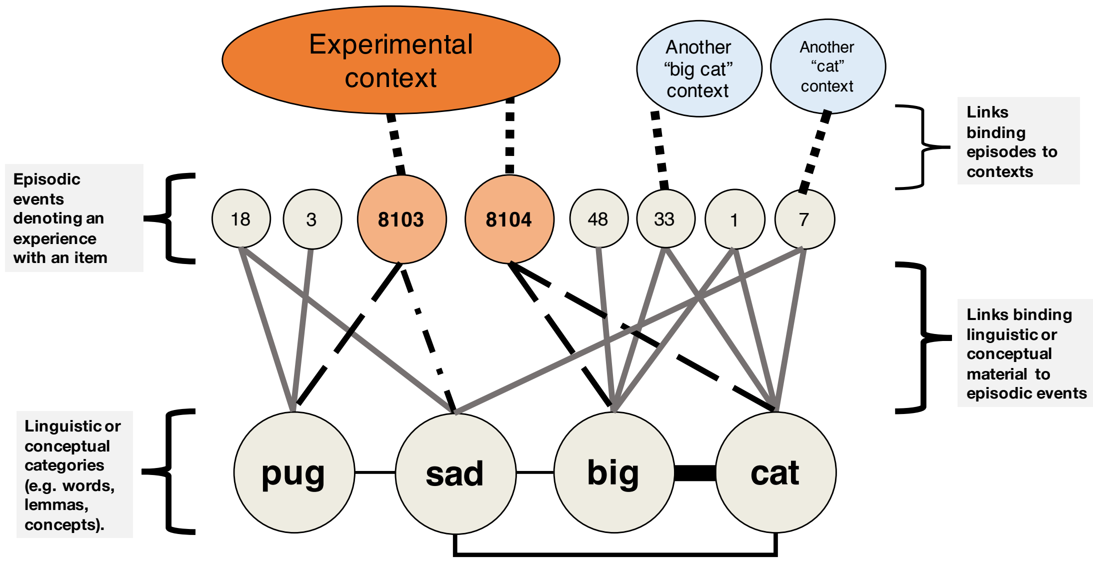

Figure 3.1: Diagram of the model of phrase frequency effects in recall and recognition. Frequent phrases are associated with more episodes and with stronger associations within the lexical-semantic network.

High frequency phrases are more likely than low frequency phrases to be completed once one word has been recalled. Because participants are capable of recalling phrases incompletely, we assume that episodes include links separately to each word in the phrase. Recall that Experiments 2a, 2b, and 3 demonstrated that phrase frequency effects arose at the level of the completion of a phrase given that recall had been initiated (that is, the q parameter value increased as phrase frequency increased). In light of these results, Figure 8 links individual words to episodes. When two words are experienced at the same time, these words attach to the same episode. This architecture allows for participants to not necessarily recall both words from a phrase. Note that there are more episodes linking *big* and *cat* together (Nodes 1, 33, and 8104) than episodes linking *sad* and *pug* (Nodes 18 and 8103). Participants must use the context to guide what items they recall: this top-down search requires also locating episodes that are associated with the experiment only and not unrelated episodes. Starting with a given context effectively eliminates all other instances of a phrase (that is, all other *big cat* episodes) during recall. Phrase frequency effects like we saw at the level of phrase completion require phrase frequency to be encoded elsewhere.

We can relate the process of retrieving both words from a phrase as being similar to spreading activation. When speakers retrieve one word, they are able to retrieve a related word more easily because words associated with previous material in long-term memory become active. In the phrase case, the next word in a phrase becomes easier to retrieve. In the architecture of this model, we represent the capacity for spreading activation between two words as solid bars connecting the words within the word layer in Figure 8. The more often two words occur together, the stronger the connection between them, and the more likely that both words will be retrieved once one has been produced.

The probability of recall of at least one word of a phrase is sometimes not affected by phrase frequency. Recall that Experiment 1 found no effect of word frequency on word recall success and we explained this by assuming that the strength of the episodic links to the words is largely independent of lexical frequency. For a non-idiomatic phrase, we assume that its episodic representation consists of links from its words to the episode. That is, there is no phrase node that is linked to the episode. Given this, we expect little effect of phrase frequency on the first stage of recall, when words are initially retrieved from the context. This is what we found in Experiments 2a and 2b, in which phrase frequency did not influence the p parameter. We note that there was an effect of phrase frequency on p in Experiment 3, though. We conclude that the frequency effect on p, unlike the consistent effect on q, comes and goes much as the effect of single-word frequency on single-word recall. It is possible that the longer study times used in

Experiment 3 allowed for the associations between words of common phrases to build the activation of the words and hence increase the strength of their links to episodic context.

Concrete phrases are easier to recall than abstract phrases. Concrete and imageable words and phrases are typically much easier to understand, produce, recognize, recall, and learn. In every experiment in this study, concreteness influenced the likelihood of the initial retrieval of a phrase (the p parameter) as well as the likelihood of the completion of a phrase given initial retrieval (the q parameter). We propose that the number of features associated with a studied word or phrase determines the strength of the link between a new episode and the item. Concrete words and phrases (e.g. *alcoholic beverages*) have a number of perceptual features that more abstract words and phrases (e.g. *psychic nephew*) do not, such as texture, color, etc. (Plaut & Shallice, 1993; Marslen-Wilson & Warren, 1994; Wiemer-Hastings & Xu, 2005; Vinson & Vigliocco, 2008; Grondin, Lupker, & McRae, 2009). These richer sensory representations make the initial retrieval of a word or phrase easier than for more abstract words and phrases.

Recognition

Any satisfactory model of phrase memory must be able to account for frequency effects in recognition memory in addition to recall. Low frequency words like *pug* are much more easily discriminated in recognition than high frequency words like *cat*. Phrase recognition differs: Chapter 1 found that participants discriminated high and low frequency *phrases* equally well, even though there was a strong bias to say that they had studied high frequency phrases like *alcoholic beverages* but not low frequency ones like *psychic nephew*. They did find that words within phrases impacted discriminability, such that participants best remembered phrases that contained low frequency words like *myocardial infarction*. In light of these results, the model must not allow for low frequency phrases to be better discriminated than high frequency phrases, but phrases with rare words should be better recognized.

How does recognition memory take place in this model? We can conceptualize recognition as the inverse of recall. Instead of going from the context to retrieving linguistic content, participants start from linguistic content in order to retrieve a context, which participants verify as part of the experiment or not. When participants read the words on a computer screen, they retrieve the episodes associated with those words (some of which overlap because of previous co-occurrence). Then, participants search within those episodes to determine whether that episode was part of the experiment.

Low frequency words are better discriminated than high frequency words. Studied low frequency words like *pug* are easier to recognize because they have fewer episodes than common words like *cat*, so participants find the experimental episode with less competition from other episodes. Unstudied low frequency words are easier to recognize because it is also easier to verify that no studied episode exists. In this respect, the model captures well-known effects captured by a number of other models (e.g. Reder et al., 2000; Hintzman, 1988; Mandler, 1980).

High frequency phrases get more "yes" responses regardless of whether they were studied or not (a bias). The bias originates from the same spreading activation-like mechanism that facilitates the completion of more common phrases in free recall. Once one word has been processed, associated words that co-occur regularly activate each other. So, once a participant has read a word like *big*, the word *cat* receives greater activation than before and is therefore easier to process. This more fluent processing leads to the illusion of the phrase having been studied – regardless of whether it was studied or not, and leads to a bias among participants to say that they have studied high frequency phrases.

High and low frequency phrases are equally well discriminated. Generally speaking, phrases are much less frequent than the words that compose them. If we assume compositional phrase representations, then recognition requires searching through episodes bound to individual words, potentially in addition to episodes bound to phrases. Following from the account in Chapter 1, we propose that the relative contribution of phrase frequency to episodic search will be much less influential than word frequency due to the existence of fewer phrase episodes, so discriminability of phrases will not be sensitive to their frequency.

Low frequency words facilitate phrase discrimination. Since the number of episodes associated with at least one word within a phrase is much larger than the number of episodes containing the whole phrase, what will experience the most interference in the search for the experimental episode will be test phrases containing high frequency words. This leads to an advantage for recognition of phrases containing uncommon words (for similar proposals, see Chapter 1, Reder et al., 2000, and Malmberg et al., 2002).

Concrete phrases are better discriminated than abstract phrases. By the same mechanism as we proposed in free recall, more concrete phrases have stronger links to an episode because they have more features. When a concrete phrase is presented during recognition, the link between that phrase and the critical episode is stronger, which leads to greater discriminability of concrete phrases.

CHAPTER 4: CONCLUSION

This dissertation examined episodic memory for words and phrases. We know from psycholinguistic research that the frequency of a phrase affects language comprehension and production above and beyond the frequency of the phrase's words. I found that phrase frequency also affects memory. It plays an influential role in phrase recognition judgments as well as phrase recall. This held true in recognition memory (Chapter 1), where common phrases were more likely to be judged as studied than unstudied, i.e. a bias with no cost to discriminability, and in free recall (Chapter 2), where high-frequency phrases were more likely to be completed than low-frequency phrases once one word had been recalled. In Chapter 3, I outlined a model that provided pathways for all of these effects to occur, while simultaneously permitting concreteness effects on recognition and recall, the word frequency mirror effect in recognition memory, and the lack of a word frequency effect on the free recall of single words. The model assumes that recall is the process of generating linguistic material given an episodic context, while recognition is the retrieval of a particular episodic context given linguistic material. I propose that the predominant driving mechanism behind phrase frequency effects is a link in long-term memory between individual words.

Altogether this work represents a novel unification of psycholinguistic and episodic memory research. The results and model are an important initial step in understanding how linguistic structures beyond the level of the individual word are represented in long-term memory.

REFERENCES

Alario, F. X., Costa, A., & Caramazza, A. (2002). Frequency effects in noun phrase production: Implications for models of lexical access. *Language and Cognitive Processes*, *17*, 299-319.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79*, 97-123.

Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, *128*, 186.

Arnon, I., & Cohen Priva, U. (2013). More than Words: The Effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, *56*, 349–371.

Arnon, I., & Priva, U. C. (2014). Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *The Mental Lexicon*, *9*, 377-400.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*, 67-82.

Baayen, R. H., Hendrix, P., & Ramscar, M. (2013). Sidestepping the combinatorial explosion: An explanation of n-gram frequency effects based on naive discriminative learning. *Language and Speech*, *56*, 329-347.

Baayen, R. H., Milin, P., Đurdević, D. F., Hendrix, P., & Marelli, M. (2010). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning.

Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 576-587.

Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). The word-frequency mirror effect in young, old, and early-stage Alzheimer's disease: Evidence for two processes in episodic recognition performance. *Journal of Memory and Language*, *46*, 199–226.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, *19*, 241-248.

Begg, I. (1972). Recall of meaningful phrases. *Journal of Verbal Learning and Verbal Behavior*, *11*, 431–439.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*, 355–387.

Bower, G. H. (1969). Chunks as interference units in free recall. *Journal of Verbal Learning and Verbal Behavior*, *8*, 610-613.

Brantz, T., & Franz, A. (2006). The google web 1T 5-gram corpus. *Linguistic Data Consortium*.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of verbal learning and verbal behavior*, *5*, 325-337.

Bybee, J. L. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*, 711-733.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*, 234-272.

Clark, S. E., & Burchett, R. E. (1994). Word frequency and list composition effects in associative recognition and recall. *Memory & Cognition*, *22*, 55–62.

Copestake, A., Lambeau, F., Villavicencio, A., Bond, F., Baldwin, T., Sag, I., & Flickinger, D. (2002). Multiword expressions: linguistic precision and reusability. In proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC 2002). pp.1-7.

Criss, A. H., Aue, W. R., & Smith, L. (2011). The effects of word frequency and context variability in cued recall. *Journal of Memory and Language*, *64*, 119–132.

Davies, M. (2008). *The corpus of contemporary American English*. BYE, Brigham Young University.

Deese, J. (1959). Influence of inter-item associative strength upon immediate free recall. *Psychological Reports*, *5*, 305–312.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *9*, 283-321.

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, *5*, 313-349.

Dell, G. S., Oppenheim, G. M., & Kittredge, A. K. (2008). Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. *Language and Cognitive Processes*, *23*, 583-608.

Dell, G.S., & Jacobs, C.L. (2015). Successful speaking: Cognitive mechanisms of adaptation in language production. In G. Hickok & S.L. Small (Eds.), *The Neurobiology of Language* (pp. 209-220). Elsevier.

Dewhurst, S., Hay, D., & Wickham, L. (2005). Distinctiveness, typicality, and recollective experience in face recognition: A principal components analysis. *Psychonomic Bulletin & Review*, *12*, 1032–1037.

Dunlap, G. L., & Dunlap, L. L. (1979). Manipulating the word frequency effect in free recall. *Memory & Cognition*, *7*, 420-425.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in second language acquisition*, *24*, 143–188.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*, 627-635.

Freebody, P., & Anderson, R. C. (1986). Serial position and rated importance in the recall of text. *Discourse Processes*, *9*, 31-36.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*, 474-496.

Garrett, M. F. (1975). The analysis of sentence production. *Psychology of Learning and Motivation*, *9*, 133-177.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*, 8–20.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*, B1-B14.

Gruenewald, P. J., & Lockhead, G. R. (1980). The free recall of category examples. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 225–240.

Harley, T. A., & Brown, H. E. (1998). What causes a tip-of-the-tongue state? Evidence for lexical neighbourhood effects in speech production. *British Journal of Psychology*, *89*, 151-174.

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, *119*, 431-440.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*, 528-551.

Horowitz, L. M., & Manelis, L. (1972). *Toward a theory of redintegrative memory: Adjective-noun phrases* (Vol. 6, pp. 193–224). Elsevier.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*, 269-299.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D., Martin, M., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *23*(5), 1217–1232.

Hulme, C., Stuart, G., Brown, G. D. A., & Morin, C. (2003). High- and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory and Language*, *49*, 500–518.

Humphreys, M. S. (1976). Frequency and recency in cued recall: Interaction of old and new learning. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(4), 413.

Janssen, N., & Barber, H. A. (2012). Phrase frequency effects in language production. *PLoS ONE*, *7*, e33202.

Jescheniak, J. D., & Levelt, W. J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 824-843.

Kahana, M. J., & Caplan, J. B. (2002). Associative asymmetry in probed recall of serial lists. *Memory & Cognition*, *30*, 841–849.

Karlsen, P., & Snodgrass, J. (2004). The word-frequency paradox for recall/recognition occurs for pictures. *Psychological Research*, *68*, 1–6.

Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology*, *25*, 463-492.

Kusyszyn, I., & Paivio, A. (1966). Transition probability, word order, and noun abstractness in the learning of adjective-noun paired associates. *Journal of Experimental Psychology*, *71*, 800-805.

Lambert, W. E., & Paivio, A. (1956). The influence of noun-adjective order on learning. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *10*, 9.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). Multiple perspectives on word production. *Behavioral and Brain Sciences*, *22*, 61-69.

Light, L. L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(3), 212–228.

Lohnas, L. J. & Kahana, M. J. (2014). Compound cuing in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 12-24.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 226.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. *Psychological Review*, *89*, 483-506.

MacLeod, C. M., & Kampe, K. E. (1996). Word frequency effects on recall, recognition, and word fragment completion tests. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*, 132–142.

Malmberg, K. J., Steyvers, M., Stephens, J. D., & Shiffrin, R. M. (2002). Feature frequency effects in recognition memory. *Memory & Cognition*, *30*, 607-613.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological review*, *87*, 252-271.

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, *101*, 653-674.

May, R. B., & Sande, G. N. (1982). Encoding expectancies and word frequency in recall and recognition. *The American Journal of Psychology*, *95*, 485-495.

Morgan, E. (2016). Generative and item-specific knowledge of language.

Morgan, E., & Levy, R. (2016). Abstract knowledge versus direct experience in processing of binomial expressions. *Cognition*, *157*, 384-402.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 676–686.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the" labor-in-vain effect.". *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 676-686.

Nozari, N., Kittredge, A. K., Dell, G. S., & Schwartz, M. F. (2010). Naming and repetition in aphasia: Steps, routes, and frequency effects. *Journal of Memory and Language*, *63*, 541-559.

Osborne, J. W., & Blackmore, D. E. (2007). Recall of adjective-noun phrases within list and prose contexts, 1–22.

Ozubko, J. D., & Joordens, S. (2007). The mixed truth about frequency effects on free recall: Effects of study list composition. *Psychonomic Bulletin & Review*, *14*, 871-876.

Paivio, A., Khan, M., & Begg, I. (2000). Concreteness and relational effects on recall of adjective-noun pairs. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *54*, 149–160.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, *10*, 377-500.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 294-320.

Rizzuto, D. S., & Kahana, M. J. (2001). An auto-associative neural network model of paired-associate learning. *Neural Computation*, *13*, 2075–2092.

Rubin, D. C. (1975). Within word structure in the tip-of-the-tongue phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *14*, 392-397.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, *21*, 168–175.

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, *115*, 893-912.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523-568.

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *37*, 776–784.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*, 302-319.

Sprenger, S. A., Levelt, W. J., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, *54*, 161-184.

Thorn, A. S. C., Gathercole, S. E., & Frankish, C. R. (2005). Redintegration and the benefits of long-term knowledge in verbal short-term memory: An evaluation of Schweickert's (1993) multinomial processing tree model. *Cognitive Psychology*, *50*, 133–158.

Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. *Perspectives on formulaic language: Acquisition and communication*, 151–173.

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Strategic search from long-term memory: An examination of semantic and autobiographical recall. *Memory*, *22*, 687–699.

Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, *40*, 183-190.

Wiemer-Hastings, K. & Xu, X. (2005). Content differences for abstract and concrete concepts. *Cognitive Science, 29*, 719-736.

Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review*, *1*, 89-106.

Yuille, J. C., Paivio, A., & Lambert, W. E. (1969). Noun and adjective imagery and order in paired-associate learning by French and English subjects. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *23*, 459–466.

Appendix A

Phrases from the Google 1T n-gram corpus (Brants & Franz, 2006) used in Experiments 1 and 2

	Phrase: Adjective	Phrase: Noun	*log_2 phrase frequency*	*log_2 adjective frequency*	*log_2 noun frequency*
Low frequency phrases	simultaneous	transduction	5.39	21.48	19.70
	downstream	subcontractors	5.42	21.57	19.98
	naughty	tot	5.64	21.88	20.14
	abandoned	arena	5.80	22.36	22.71
	accompanying	visions	6.33	22.31	20.91
	packaged	hunts	6.37	21.72	19.43
	chrome	throttle	6.50	21.53	20.22
	optimum	staining	6.50	21.69	20.45
	flaming	bounds	6.55	19.67	21.65
	predominant	organ	6.70	20.15	22.39
	psychic	nephew	6.85	21.10	20.43
	transgenic	allele	6.91	20.17	19.88
	inhaled	compounds	7.04	19.60	22.64
	programmable	fuse	7.20	20.82	20.55
	sleek	fleece	7.79	20.91	20.68
	piercing	headache	8.57	21.04	21.47
	metropolitan	zones	9.09	21.69	22.61
	decadent	era	9.19	19.22	23.28
	commanding	brigade	9.29	20.23	19.95
	distinct	affinity	9.38	23.20	21.07
	routine	expressions	9.48	23.32	22.56
	untreated	asthma	9.51	20.18	22.02
	painful	consciousness	9.66	22.27	22.50
	tangled	headset	9.74	19.34	21.38
	intense	cultivation	9.79	22.76	20.93
	perennial	grasslands	10.29	20.41	19.034
High frequency phrases	thick	bundles	10.30	23.50	20.49
	vibrant	acidity	10.80	21.61	19.28
	polynomial	curves	11.04	21.11	22.09
	cherished	traditions	11.97	19.88	22.31
	passionate	embrace	13.18	21.58	21.71
	accumulated	surplus	13.24	21.61	22.18
	conditional	expectation	14.97	21.83	21.80
	relentless	pursuit	15.13	19.84	21.84
	unsecured	tenant	15.32	21.64	21.81
	roman	numerals	15.56	20.28	19.25
	interior	decoration	16.06	23.48	21.41

contaminated	soils	16.35	21.81	21.83
undue	hardship	16.94	20.31	20.60
outer	shell	17.35	22.83	23.43
dining	hall	17.55	23.44	23.09
mashed	potatoes	18.34	19.37	21.71
respiratory	tract	18.59	22.01	21.93
cystic	fibrosis	18.67	19.37	19.85
cerebral	palsy	18.73	20.98	19.39
monoclonal	antibody	18.75	19.99	22.03
bald	eagle	18.82	22.00	21.54
nitric	oxide	19.30	19.75	21.74
myocardial	infarction	19.42	20.37	19.93
coronary	artery	19.53	21.29	21.35
alcoholic	beverages	19.56	21.34	21.55
rheumatoid	arthritis	19.65	19.93	21.79

Appendix B

Phrases derived from COCA (Davies, 2008-) used in Experiment 2.

	Phrase				
	Adjective	Noun	*log$_2$ phrase frequency*	*log$_2$ adjective frequency*	*log$_2$ noun frequency*
Low frequency phrases	poor	credit	2.32	12.82	12.79
	southern	food	2.81	12.36	13.81
	fantastic	panel	2.81	11.18	12.19
	nice	hair	3	13.37	12.69
	incredible	pain	3.17	12.28	12.75
	safe	space	3.17	12.81	13.04
	available	flight	3.17	12.78	12.42
	controversial	statement	3.32	11.71	13.19
	violent	weather	3.32	12.04	12.41
	similar	incident	3.46	12.45	12
	particular	church	3.46	13.12	13.35
	local	airport	3.58	13.57	12.09
	open	relationship	3.7	12.77	13.59
	heavy	heart	3.7	12.09	13.86
	likely	suspect	3.7	12.64	11.35
	impossible	dream	3.91	11.87	12.18
	wonderful	trip	4	13.51	12.58
	British	actor	4	12.66	12.56
	serious	nature	4.09	13.83	12.44
	major	bank	4.09	13.94	12.61
	crazy	talk	4.09	12.45	13.43
	sad	truth	4.17	12.05	13.53
	successful	mission	4.17	12.66	12.84
	simple	rule	4.17	12.91	11.94
	global	recession	4.17	12.14	11.8
	angry	crowd	4.25	12.62	11.97
	late	term	4.25	12.53	12.82
	guilty	pleasure	4.25	13	12.2
	normal	behavior	4.39	12.74	12.43
	fresh	blood	4.39	12.05	13.2
	strong	opinion	4.46	13.73	13.05
	healthy	weight	4.46	12.03	12.25
	super	model	4.52	11.57	11.9
	funny	feeling	4.58	12.76	12.47
	necessary	step	4.58	12.43	12.31
	positive	test	4.58	12.63	12.57

	lucky	break	4.64	11.82	14.74
	current	governor	4.64	12.59	13.46
	actual	cost	4.81	11.87	12.55
	easy	solution	4.86	12.97	12.16
	civil	union	4.86	13.19	13.19
	horrible	mistake	4.86	11.68	12.58
	fair	deal	4.91	12.44	13.61
	international	agreement	4.91	13.77	12.94
	clear	winner	4.91	13.42	11.68
	famous	speech	4.91	12.52	13.38
	effective	treatment	4.95	12.23	12.71
	white	neighborhood	5	13.57	12.298
	sexual	act	5	12.85	12.73
	legal	strategy	5.04	13.46	12.46
	senior	officer	5.09	12.89	12.95
	military	background	5.13	14.14	12.05
	quick	action	5.17	12.63	13.62
	full	picture	5.21	13.48	13.45
	short	film	5.25	12.79	13.47
High frequency phrases	liberal	agenda	5.29	11.94	12.31
	dangerous	drug	5.32	12.88	13.72
	afghan	border	5.43	10.87	12.6
	commercial	success	5.46	14.21	12.75
	physical	violence	5.49	12.18	13.43
	emotional	response	5.55	12.07	12.81
	innocent	victim	5.58	11.92	12.42
	terrible	accident	5.64	12.8	12.21
	prime	example	5.64	12.87	12.67
	Iraqi	freedom	5.67	13.51	12.7
	extraordinary	amount	5.75	12.1	13.18
	specific	threat	5.81	12.55	12.9
	amazing	experience	5.88	12.83	13.39
	beautiful	song	5.93	13.34	13.08
	private	plane	5.95	13.32	12.9
	certain	type	5.98	13.88	12.9
	personal	choice	6	13.56	13.13
	social	network	6	13.57	12.5
	entire	industry	6.02	13.16	13.54
	fine	art	6.09	13.23	12.56
	powerful	message	6.11	12.58	13.63
	independent	investigation	6.27	12.44	13.56
	smart	move	6.3	11.96	11.74

significant	progress	6.34	12.59	12.24
main	course	6.38	12.6	12.64
correct	answer	6.39	12.63	13.48
supreme	leader	6.44	13.18	13.57
enormous	pressure	6.58	12.18	13.12
red	tape	6.74	12.13	12.8
financial	reform	6.88	12.91	13.22
tough	love	6.93	13.58	13.44
perfect	storm	7	12.58	12.44
religious	right	7.06	12.46	13.09
close	attention	7.17	12.72	13.57
dead	heat	7.26	13.37	11.55
hot	seat	7.31	12.76	12.16
single	parent	7.46	13.27	11.98
critical	condition	7.5	12.51	11.81
low	income	7.53	12.03	12.42
recent	study	7.58	13.05	12.68
early	age	7.88	13.22	13.56
wrong	direction	8.06	12.78	12.4
central	park	8.23	12.74	12.16
popular	vote	8.4	12.73	13.44
congressional	budget	8.56	12.36	13.7
regular	basis	8.75	12.07	12.47
common	ground	8.84	12.72	13.3
free	market	8.97	13.58	13.5
natural	gas	9	12.18	12.68
nuclear	weapon	9.02	13.47	11.68
illegal	immigration	9.18	12.54	12.06
gay	marriage	9.25	12.41	12.96
economic	growth	9.9	13.92	12.52
foreign	language	10.6	13.82	12.86
presidential	candidate	11.24	13.49	13.45
middle	class	11.27	13.13	13.16
grand	jury	11.32	12.18	13.72

Appendix C

Phrases and nouns derived from Balota et al. (2002) and used in Experiments 3a and 3b.

	Phrase Adjective	Noun	*log_2 phrase frequency*	*log_2 adjective frequency*	*log_2 noun frequency*
High frequency nouns	adjacent	nation	7.11	23.04	24.74
	ambitious	library	7.81	21.55	25.33
	artificial	home	7.70	22.48	28.37
	awesome	valley	7.09	23.29	22.64
	beloved	chicken	7.59	21.78	23.16
	beneficial	sun	7.07	22.59	24.38
	biological	garden	7.23	23.43	24.25
	bold	rose	7.65	22.80	23.67
	burning	palace	7.99	23.25	21.54
	cooling	floor	7.56	22.63	24.99
	cycling	town	7.75	21.85	25.42
	destructive	baby	7.22	21.43	25.17
	downstream	field	7.42	21.57	26.41
	emerging	road	7.95	23.03	25.36
	endless	cloud	8.08	21.96	22.42
	engaged	father	7.21	23.49	24.92
	failing	hotel	6.90	22.46	26.58
	gentle	snake	6.85	22.27	21.56
	governing	village	7.84	22.74	23.99
	grounded	world	7.29	21.66	27.64
	hanging	dress	7.28	22.77	23.68
	hazardous	car	7.42	22.70	26.68
	inspiring	college	6.91	21.46	25.39
	instructional	kitchen	6.97	22.16	24.17
	insured	truck	7.27	22.28	23.84
	invisible	mouth	7.58	22.01	24.18
	jumping	cow	7.43	21.72	22.02
	literary	radio	8.00	22.57	25.28
	magnificent	plane	7.93	21.93	23.72
	metallic	wheel	7.12	21.43	23.71
	patented	bottle	7.83	21.44	23.31
	premature	tree	6.96	21.42	25.01
	provincial	street	6.85	22.33	25.01
	refurbished	engine	7.88	21.41	25.34
	rejected	picture	7.09	22.92	26.16
	rolled	bread	8.08	22.32	22.98
	specialized	pool	8.02	22.90	24.77
	stainless	key	6.98	22.67	26.37

	sterling	cup	7.42	22.11	23.99
	sticky	book	7.02	21.45	27.28
	stolen	jacket	7.24	22.35	22.66
	striking	beach	7.77	22.37	24.62
	surfing	market	7.26	21.95	26.66
	surprised	cat	7.48	23.21	24.09
	teenage	king	7.25	23.18	23.74
	tough	bear	7.12	23.38	23.88
	toxic	stream	7.09	22.50	24.90
	versatile	ball	7.15	21.83	24.45
Low frequency nouns	adjustable	anvil	6.84	22.45	18.11
	bald	vulture	6.87	22.00	17.52
	bitter	pecan	7.97	21.76	18.21
	blind	owl	7.68	23.27	20.54
	brilliant	sleuth	6.84	22.83	17.05
	circular	parasol	7.24	22.01	16.87
	complementary	valet	7.42	21.64	19.13
	copper	altar	8.05	22.70	20.77
	crowded	isle	7.14	21.43	19.24
	cute	otter	7.02	23.40	18.49
	deadly	dungeon	7.82	21.83	19.65
	decorative	gourd	7.86	21.75	17.90
	delicate	sequin	7.61	21.71	18.26
	dried	eel	7.52	22.14	18.76
	elegant	lily	7.22	22.67	20.14
	expanding	cavern	7.53	22.87	19.75
	extraordinary	gem	7.04	22.64	21.17
	fake	cobra	6.80	22.74	19.41
	fancy	loft	7.30	22.21	20.27
	golden	plum	7.82	22.92	19.75
	grey	bonnet	7.90	22.52	18.97
	handsome	wizard	7.81	21.44	21.64
	indigenous	spa	7.19	22.12	22.52
	lively	lass	6.89	21.54	18.45
	miniature	tripod	7.97	21.44	20.31
	nasty	beggar	6.96	22.28	18.31
	occasional	jaguar	6.76	22.26	18.88
	offshore	wharf	6.82	22.33	18.77
	ordinary	flea	7.43	23.43	20.21
	polished	flask	7.76	21.66	19.14
	portable	keg	6.85	23.47	18.24
	relaxing	harp	7.40	21.91	19.91
	robust	vine	7.07	22.52	20.15
	sacred	urn	7.96	22.06	19.83
	shallow	crevice	7.18	21.86	17.41

silly	dwarf	7.38	22.27	20.35
slim	vase	7.93	21.89	20.57
spinning	galaxy	7.35	21.40	21.21
stuffed	boar	7.24	21.40	18.68
stylish	yacht	7.45	22.88	20.97
tan	tunic	7.28	21.59	18.52
thin	tablet	7.31	23.42	21.77
tropical	olive	7.80	22.66	21.80
vintage	banjo	7.99	22.98	19.51
wooden	silo	7.01	22.76	18.32
yearly	monsoon	7.41	21.83	19.09

Appendix D

Materials derived from Chapter 1 in recall Experiments 2a and 2b.

adjective	noun	log adjective frequency	log noun frequency	log phrase frequency	imageability	compositionality
effective	treatment	10.22	10.61	2.32	10.38	10.48
impossible	dream	10.01	10.55	2.32	10.0	10
open	relationship	10.92	11.80	2.32	10.52	9.62
poor	credit	10.74	11.21	2.32	10.17	9.52
sad	truth	10.36	11.67	2.32	10.41	9.97
serious	nature	11.94	10.33	2.32	10.79	9.83
similar	incident	10.61	10.19	2.32	10.1	10.17
fair	deal	10.46	11.95	2.58	10.17	10.18
funny	feeling	11.21	10.23	2.58	10.21	8.86
heavy	heart	10.18	11.92	2.58	10.31	8.76
major	bank	11.89	10.95	2.58	10.04	9.44
physical	violence	10.31	11.34	2.58	10.59	10.38
british	actor	10.55	10.65	2.81	10.31	10.41
necessary	step	10.15	10.39	2.81	10.21	10.0
normal	behavior	10.76	10.54	2.81	10.31	10.14
positive	test	10.75	10.65	2.81	10.03	9.41
safe	space	10.96	11.08	2.81	10.15	10.21
successful	mission	10.71	10.92	2.81	10.15	10.31
violent	weather	10.01	11.16	2.81	10.55	10.1
actual	cost	10.16	10.77	3	10.34	9.59
available	flight	10.64	10.59	3	10.38	10.31
easy	solution	11.05	10.26	3	10.38	10.34
fresh	blood	10.68	11.29	3	10.31	9.55
iraqi	freedom	10.97	10.6	3	10	8.9

adjective	noun	log adjective frequency	log noun frequency	log phrase frequency	imageability	compositionality
quick	action	10.9	11.29	3	10.07	10.07
senior	officer	11.06	10.81	3	10.24	9.48
white	neighborhood	11.69	10.34	3	10.41	9.34
international	agreement	11.62	10.35	3.17	10	10.07
full	picture	11.59	11.64	3.32	10.55	8.89
likely	suspect	10.8	10.05	3.32	10.38	9.69
lucky	break	10.17	11.64	3.32	10.07	9.21
strong	opinion	11.85	10.98	3.32	10.34	9.97
terrible	accident	10.87	10.6	3.32	10.59	10.07
clear	winner	11.55	10.08	3.46	10.14	9.17
current	governor	10.82	11.81	3.46	10.38	10.24
fine	art	11.35	10.01	3.46	10.11	8.93
military	background	11.84	10.04	3.46	10.24	9.38
super	model	10.41	10.29	3.46	10.69	8.72
emotional	response	10.35	11	3.58	10.41	10.21
horrible	mistake	10.13	10.67	3.58	10.52	10.31
sexual	act	10.56	10.9	3.58	10.35	10.55
short	film	10.98	11.50	3.58	10.48	10.31
commercial	success	10.38	10.79	3.7	10.17	9.07
global	recession	10.89	10.27	3.7	10.14	10.18
healthy	weight	10.35	10.61	3.81	10.41	9.97
guilty	pleasure	11.19	10.36	4	10.31	9.45
innocent	victim	10.04	10.63	4.17	10.24	10.34
extraordinary	amount	10.19	11.3	4.39	10.07	10.03
personal	choice	11.55	11.16	4.39	10.24	10.1
independent	investigation	10.15	11.74	4.58	9.97	9.79
beautiful	song	11.86	11.33	4.64	10.32	10.41

adjective	noun	log adjective frequency	log noun frequency	log phrase frequency	imageability	compositionality
significant	progress	10.87	10.45	4.75	10.24	10.03
amazing	experience	11.57	11.61	4.81	10.34	10.52
correct	answer	10.22	11.47	4.81	10.59	10.59
enormous	pressure	10.15	11.21	4.81	10.29	9.72
powerful	message	10.73	11.84	4.81	10.31	10.21
private	plane	11.51	11.19	4.91	10.69	10.38
single	parent	11.57	10.07	5.17	10.69	9.97
close	attention	10.91	11.71	5.25	10.34	8.79
main	course	10.58	10.88	5.46	10.59	8.9
recent	study	11.09	10.72	5.58	10.34	10.21
tough	love	11.89	11.72	5.61	10.1	9.69
early	age	11.72	11.62	5.7	10.59	9.69
low	income	10.1	10.61	5.91	10.52	10.1
social	network	11.61	10.88	5.95	10.45	9.31
supreme	leader	10.93	11.5	6.07	10.31	9.9
hot	seat	11.14	10.69	6.25	10	8.31
critical	condition	10.67	10.03	6.3	10.38	10.17
wrong	direction	10.98	10.65	6.8	10.41	10.31
popular	vote	10.82	11.43	6.88	10.34	9.24
regular	basis	10.29	10.06	6.89	10.21	9.28
common	ground	10.80	11.64	7.3	10.21	8.71

Appendix E: IRB Letter

UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN

Office of the Vice Chancellor for Research

Office for the Protection of Research Subjects
528 East Green Street
Suite 203
Champaign, IL 61820

April 28, 2016

Gary Dell
Psychology
2117 Beckman
405 North Mathews Avenue
Urbana, Il 61801

RE: *The role of information structure on word learning*
IRB Protocol Number: 13380

Dear Dr. Dell:

This letter authorizes the use of human subjects in your continuing project entitled *The role of information structure on word learning*. The University of Illinois at Urbana-Champaign Institutional Review Board (IRB) approved the protocol as described in your IRB application, by expedited continuing review. The expiration date for this protocol, IRB number 13380, is 04/27/2017. The risk designation applied to your project is *no more than minimal risk*.

Copies of the attached date-stamped consent form(s) must be used in obtaining informed consent. If there is a need to revise or alter the consent form(s), please submit the revised form(s) for IRB review, approval, and date-stamping prior to use.

Under applicable regulations, no changes to procedures involving human subjects may be made without prior IRB review and approval. The regulations also require that you promptly notify the IRB of any problems involving human subjects, including unanticipated side effects, adverse reactions, and any injuries or complications that arise during the project.

If you have any questions about the IRB process, or if you need assistance at any time, please feel free to contact me at the OPRS office, or visit our Web site at http://oprs.research.illinois.edu.

Sincerely,

LeaAnn Carson

LeaAnn Carson, MS
OPRS Specialist

Attachment(s)

c: Cassandra Jacobs

U of Illinois at Urbana-Champaign • IORG0000014 • FWA #00008584
telephone (217) 333-2670 • *fax* (217) 333-0405 • *email* IRB@illinois.edu